

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 15:30:32 Search time 48 Seconds  
(without alignments)  
4296.941 Million cell updates/sec

Title: US-10-046-433-40

Perfect score: 1001  
Sequence: 1 MAEPGSHHLSARVRGRTER.....LGRSNHLPRGLMDLTQCR 1001

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 150 summaries

Database:

SPTREMBL.21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_nbc.\*  
8: sp\_organelle.\*  
9: sp\_plant.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriopl.\*  
17: sp\_archaeop.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	554	55.3	580	4 Q9P2M2	Q9P2M2 homo sapien
2	91	9.1	300	11 Q8R215	Q8R215 mus musculu
3	15	1.5	493	4 Q96DP2	Q96DP2 homo sapien
4	8	0.8	97	10 Q42206	Q42206 arabidopsis
5	8	0.8	133	12 Q9PXR8	Q9PXR8 hepatitis c
6	8	0.8	248	10 Q9C5R4	Q9C5R4 hepatitis c
7	8	0.8	302	12 Q8V7G7	Q8V7G7 tt virus. o
8	8	0.8	353	10 Q42584	Q42584 arabidopsis
9	8	0.8	353	10 Q9SM08	Q9SM08 arabidopsis
10	8	0.8	430	16 Q9A0D7	Q9A0D7 streptococc
11	8	0.8	444	16 Q921L4	Q921L4 rickettsia
12	8	0.8	695	16 Q92891	Q92891 chlamydia p
13	8	0.8	695	16 Q93S58	Q93S58 chlamydia p
14	8	0.8	744	3 Q42958	Q42958 schizosach
15	7	0.7	77	5 Q95XN6	Q95XN6 caenorhabdi
16	7	0.7	79	5 Q9W2P5	Q9W2P5 drosophila

17	7	0.7	91	10 Q22823	Q22823 arabidopsis
18	7	0.7	91	10 Q9M1Z3	Q9M1Z3 arabidopsis
19	7	0.7	93	2 Q9F1P2	Q9F1P2 enterococcu
20	7	0.7	100	10 Q9XJPI	Q9XJPI aureococcu
21	7	0.7	106	16 Q97RY2	Q97RY2 streptococ
22	7	0.7	115	5 Q9N7S4	Q9N7S4 leishmania
23	7	0.7	116	10 Q9M280	Q9M280 arabidopsis
24	7	0.7	122	2 Q9RAN1	Q9RAN1 methylobact
25	7	0.7	124	5 Q94045	Q94045 caenorhabdi
26	7	0.7	132	5 Q96404	Q96404 plasmodium
27	7	0.7	134	17 Q97UE3	Q97UE3 sulfobact
28	7	0.7	138	2 Q9EUI6	Q9EUI6 sulfobact
29	7	0.7	141	2 Q95567	Q95567 mycobacteri
30	7	0.7	159	17 Q9HRJ4	Q9HRJ4 halobacteri
31	7	0.7	163	5 Q9U732	Q9U732 cryptospori
32	7	0.7	163	5 Q9GQD9	Q9GQD9 cryptospori
33	7	0.7	165	10 Q8S448	Q8S448 mentha long
34	7	0.7	165	10 Q8S3X6	Q8S3X6 mentha long
35	7	0.7	167	5 Q9U733	Q9U733 cryptospori
36	7	0.7	167	5 Q9GQEI	Q9GQEI cryptospori
37	7	0.7	167	5 Q9GQCI	Q9GQCI cryptospori
38	7	0.7	173	10 Q9S709	Q9S709 rhodobacter
39	7	0.7	188	3 Q9UTV4	Q9UTV4 streptomyce
40	7	0.7	192	4 Q9Y329	Q9Y329 oryza sativ
41	7	0.7	197	16 Q9KMR8	Q9KMR8 oryza sativ
42	7	0.7	210	2 Q33570	Q33570 rhodobacter
43	7	0.7	211	16 Q931V4	Q931V4 streptomyce
44	7	0.7	215	16 Q8Z1P0	Q8Z1P0 yersinia pe
45	7	0.7	217	5 Q9VVA5	Q9VVA5 drosophila
46	7	0.7	223	10 Q941R6	Q941R6 oryza sativ
47	7	0.7	223	16 Q9ZK02	Q9ZK02 helicobacte
48	7	0.7	224	2 Q9S0D9	Q9S0D9 borrelia bu
49	7	0.7	224	16 Q93S03	Q93S03 streptomyce
50	7	0.7	226	3 Q42826	Q42826 schizosach
51	7	0.7	226	16 Q25822	Q25822 helicobacte
52	7	0.7	241	8 Q9WP30	Q9WP30 forelius ch
53	7	0.7	246	5 Q8STU3	Q8STU3 encephalito
54	7	0.7	247	2 Q07821	Q07821 paracoccus
55	7	0.7	248	5 Q9N9P6	Q9N9P6 leishmania
56	7	0.7	258	3 Q9HFM6	Q9HFM6 penicillium
57	7	0.7	267	17 Q8ZXA9	Q8ZXA9 pyrobaculum
58	7	0.7	275	16 Q8STF1	Q8STF1 encephalito
59	7	0.7	275	16 Q92T29	Q92T29 rhizobium m
60	7	0.7	282	16 Q8YSU3	Q8YSU3 anabaena sp
61	7	0.7	284	16 Q8U7C6	Q8U7C6 agrobacteri
62	7	0.7	285	3 Q9P4S1	Q9P4S1 penicillium
63	7	0.7	294	5 Q94295	Q94295 caenorhabdi
64	7	0.7	294	16 Q9AB93	Q9AB93 caulobacter
65	7	0.7	297	2 Q9KIF0	Q9KIF0 streptomyce
66	7	0.7	300	16 Q98BC8	Q98BC8 rhizobium l
67	7	0.7	300	16 Q8ZLB0	Q8ZLB0 yersinia pe
68	7	0.7	300	16 Q8ZAB1	Q8ZAB1 yersinia pe
69	7	0.7	300	16 Q8Z298	Q8Z298 salmonella
70	7	0.7	300	16 Q8UC61	Q8UC61 salmonella
71	7	0.7	300	16 Q92LJ5	Q92LJ5 agrobacteri
72	7	0.7	300	16 Q8XJBS	Q8XJBS rhizobium m
73	7	0.7	303	16 Q9Z799	Q9Z799 brucella me
74	7	0.7	305	16 Q9Z799	Q9Z799 chlamydia p
75	7	0.7	306	2 Q96567	Q96567 vibrio angu
76	7	0.7	309	4 Q96SE2	Q96SE2 homo sapien
77	7	0.7	310	16 Q8Z180	Q8Z180 yersinia pe
78	7	0.7	312	16 Q9EWZ9	Q9EWZ9 streptomyce
79	7	0.7	315	16 Q97FUI	Q97FUI clostridium
80	7	0.7	317	16 Q981S2	Q981S2 rhizobium l
81	7	0.7	318	10 Q43055	Q43055 populus kit
82	7	0.7	325	12 Q91E43	Q91E43 phocid herp
83	7	0.7	327	10 Q9CAX9	Q9CAX9 arabidopsis
84	7	0.7	327	16 Q9PCG1	Q9PCG1 ureaplasma
85	7	0.7	336	16 Q97KX8	Q97KX8 clostridium
86	7	0.7	347	10 Q8RYP3	Q8RYP3 gossypium h
87	7	0.7	350	16 Q8Z0T2	Q8Z0T2 salmonella
88	7	0.7	350	16 Q8Z8B5	Q8Z8B5 salmonella
89	7	0.7	357	5 Q8SUY1	Q8SUY1 encephalito

90	7	0.7	360	16	Q92HR6	Q92hr6 rickettsia
91	7	0.7	361	16	Q9CN76	Q9cn76 pasteurella
92	7	0.7	364	16	Q9SKA0	Q9ska0 macaca fasc
93	7	0.7	364	16	Q9FCI6	Q9fci6 streptomyce
94	7	0.7	371	5	P91984	P91984 caenorhabdi
95	7	0.7	371	16	Q9CBV6	Q9cbv6 mycobacteri
96	7	0.7	371	16	Q9CBV6	Q9cbv6 mycobacteri
97	7	0.7	386	11	Q91WP3	Q91wp3 mus musculi
98	7	0.7	386	11	Q91WP3	Q91wp3 mus musculi
99	7	0.7	387	16	Q9Y4S3	Q9y4s3 homo sapien
100	7	0.7	389	16	Q9PC12	Q9pc12 xylella fas
101	7	0.7	391	16	Q9W607	Q9w607 staphylococ
102	7	0.7	392	11	Q9P883	Q9p883 agaricus bi
103	7	0.7	392	11	Q9CY58	Q9cy58 mus musculi
104	7	0.7	392	11	Q9VHU3	Q9vhu3 rattus norv
105	7	0.7	393	4	Q9Y367	Q9y367 homo sapien
106	7	0.7	402	4	Q9EUM4	Q9eum4 homo sapien
107	7	0.7	402	4	Q8WUHO	Q8wuh0 homo sapien
108	7	0.7	404	16	Q9AK93	Q9ak93 streptomyce
109	7	0.7	407	11	Q9DBY6	Q9dby6 mus musculi
110	7	0.7	408	4	Q9ERTY3	Q9erty3 homo sapien
111	7	0.7	409	9	Q64275	Q64275 lactococcus
112	7	0.7	410	4	Q96PC2	Q96pc2 homo sapien
113	7	0.7	413	2	Q931F7	Q931f7 ptiophilobac
114	7	0.7	419	16	Q8XSW3	Q8xsw3 talictonia s
115	7	0.7	422	4	Q9BWF1	Q9bwf1 homo sapien
116	7	0.7	422	16	Q9CMS4	Q9cms4 pasteurella
117	7	0.7	423	16	Q8ZD48	Q8zda8 yersinia pe
118	7	0.7	431	16	Q8YSW8	Q8ysw8 anabaena sp
119	7	0.7	434	16	Q98336	Q98336 guillardi
120	7	0.7	436	16	Q9XAZ7	Q9xaz7 streptomyce
121	7	0.7	438	5	Q9N948	Q9n948 trypanosoma
122	7	0.7	445	5	Q9N811	Q9n811 trypanosoma
123	7	0.7	446	5	Q9VJ92	Q9vj92 diosiphila
124	7	0.7	450	2	Q55109	Q55109 synecococc
125	7	0.7	450	10	Q9SLB9	Q9slb9 arabidopsis
126	7	0.7	450	12	Q91S68	Q91s68 planaria as
127	7	0.7	451	10	Q9MLX6	Q9mlx6 arabidopsis
128	7	0.7	452	10	Q9LWZ3	Q9lwz3 oryza sativ
129	7	0.7	468	8	Q32863	Q32863 pedionomas
130	7	0.7	471	16	Q8XZY7	Q8xzy7 ralsconia s
131	7	0.7	474	4	Q96AF8	Q96af8 homo sapien
132	7	0.7	482	17	Q58773	Q58773 methanococc
133	7	0.7	495	5	Q81988	Q81988 listeria mo
134	7	0.7	508	5	Q9UAR0	Q9uar0 caenorhabdi
135	7	0.7	515	10	Q82588	Q82588 arabidopsis
136	7	0.7	515	10	Q9CSM7	Q9csm7 arabidopsis
137	7	0.7	515	12	Q64784	Q64784 avian adeno
138	7	0.7	521	16	Q8XP53	Q8xp53 clostridium
139	7	0.7	524	4	Q94818	Q94818 homo sapien
140	7	0.7	530	10	Q9SG67	Q9sg67 lycopersico
141	7	0.7	530	10	Q9SG66	Q9sg66 lycopersico
142	7	0.7	530	10	Q82811	Q82811 arabidopsis
143	7	0.7	530	16	Q99XAX3	Q99xax3 staphylococ
144	7	0.7	531	10	Q8W071	Q8w071 oryza sativ
145	7	0.7	538	4	Q9H1W6	Q9h1w6 homo sapien
146	7	0.7	538	11	Q8R562	Q8r562 mus musculi
147	7	0.7	549	3	Q93956	Q93956 emeticella
148	7	0.7	558	10	Q9FM20	Q9fm20 arabidopsis
149	7	0.7	558	12	Q36296	Q36296 venezuelan
150	7	0.7	558	12	Q36293	Q36293 venezuelan

## ALIGNMENTS

## RESULT 1

Q9P2M2 PRELIMINARY: PRT: 580 AA.  
 AC Q9P2M2, 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE KIAA1324 protein (Fragment).  
 GN KIAA1324.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RX MEDLINE=20181126; PubMed=10718198;  
 RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVI.  
 RT The complete sequences of 150 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 7:65-73(2000).  
 DR EMBL; AB037745; BAA92562.1; -  
 FT NON TER  
 SQ SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;

Query Match 55.3%; Score 554; DB 4; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 554; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	425	WMNTLPMTMETVYLSGINFEYKMGTEWVAGDHIYTAAGASDNDPMILTVVPGFRPQS	484
Db	1	WMNTLPMTMETVYLSGINFEYKMGTEWVAGDHIYTAAGASDNDPMILTVVPGFRPQS	60
Qy	485	VMADTENKEVARITVEETLCSVNCLEIFMVGVNSRTNTPVETWKSXGKOSTYIIEN	544
Db	61	VMADTENKEVARITVEETLCSVNCLEIFMVGVNSRTNTPVETWKSXGKOSTYIIEN	120
Qy	545	TTTSEPTAFORTTHERASRYTNDVAKIYSINTVWNGVASYCRPCALASDVSSCTS	604
Db	121	TTTSEPTAFORTTHERASRYTNDVAKIYSINTVWNGVASYCRPCALASDVSSCTS	180
Qy	605	CPAGYIIDRDSGTCHSCPNTILKAHQYGVQACVPCGPKTKNKHLSLCYNDCTSRNT	664
Db	181	CPAGYIIDRDSGTCHSCPNTILKAHQYGVQACVPCGPKTKNKHLSLCYNDCTSRNT	240
Qy	665	PTTFNNVNFSLANTVTLAGSPSTSGKLYFHHFTLSLGNQGRKMSVCTDVTDLRIP	724
Db	241	PTTFNNVNFSLANTVTLAGSPSTSGKLYFHHFTLSLGNQGRKMSVCTDVTDLRIP	300
Qy	725	EGESGFSKSIYAVYCOAVIIPPEVTGYKAGVSGQVSLADRLIGVTTMTLDTGTSFAL	784
Db	301	EGESGFSKSIYAVYCOAVIIPPEVTGYKAGVSGQVSLADRLIGVTTMTLDTGTSFAL	360
Qy	785	PHLESIGIPDVIFFYRSNDVTQSCSGRSTTRRCSPQKTVPSLILPCTSGDGTGDC	844
Db	361	PHLESIGIPDVIFFYRSNDVTQSCSGRSTTRRCSPQKTVPSLILPCTSGDGTGDC	420
Qy	845	NHFMESAAACPLCSVDYHAIVSSCVAGIOKTYYMRBPKLCSGSIPLBQRTVICT	904
Db	421	NHFMESAAACPLCSVDYHAIVSSCVAGIOKTYYMRBPKLCSGSIPLBQRTVICT	480
Qy	905	IDFMILKVISAGTCTAAILTVLTCYFMKKNQKLEYKSKLVNMTLTKDCDLPADSCAIM	964
Db	481	IDFMILKVISAGTCTAAILTVLTCYFMKKNQKLEYKSKLVNMTLTKDCDLPADSCAIM	540
Qy	965	EGEDVEDDLIFTSK 978	
Db	541	EGEDVEDDLIFTSK 554	

## RESULT 2

Q8R215 PRELIMINARY: PRT: 300 AA.  
 AC Q8R215, 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DE Hypothetical 32.6 kDa protein (Fragment).  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COLON;  
 RA Strausberg R.;  
 RL Submitted (FE8-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC022655; AA022655.1; -  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;

Query Match  
 Best Local Similarity 9.1%; Score 91; DB 11; Length 300;  
 Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 882 WREPLCSGSLPFRQYTKTIDFWLKVGISAGCTAILLTVLTCYFWKKQKLEKYY 941  
 DB 169 WREPLCSGSLPFRQYTKTIDFWLKVGISAGCTAILLTVLTCYFWKKQKLEKYY 228

OY 942 SKLVNATLKDCDLPADSCAIMEGEDEDD 972  
 DB 229 SKLVNATLKDCDLPADSCAIMEGEDEDD 259

## RESULT 3

OY 096DP2 PRELIMINARY; PRT; 493 AA.

AC 096DP2;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
 DE CDNA FLJ13140 f1s, clone MESAN100035, weakly similar to major  
 DE surface-labeled trophozoite antigen precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RA Tashiro H., Yamazaki M., Matanabe K., Kumagai A., Itakura S.,  
 RA Fukuzumi Y., Fujimori Y., Komiya M., Sugiyama T., Ito R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hiro Y., Saito K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Matsuo M.,  
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,  
 RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Maeno Y., Nagai K.,  
 RA Iisaga T.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK055502; BAB71041.1; -  
 DR InterPro; IPR001881; EGF CA.  
 DR SWARK; SM00179; EGF CA; 1.  
 SQ SEQUENCE 493 AA; 54585 MW; 0AEBB2C19709B211 CRC64;

Query Match  
 Best Local Similarity 1.5%; Score 15; DB 4; Length 493;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 268 GVAYTSECFCKPCT 282  
 DB 43 GVAYTSECFCKPCT 57

## RESULT 4

OY 042206 PRELIMINARY; PRT; 97 AA.

AC 042206;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE Peroxidase (Fragment).

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA; TISSUE=SEEDLING;  
 RA Desprez T., Amselem J., Chiapello H., Caboche M., Hofte H.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z29133; CA82392.1; -  
 DR HSSP; P00433; ZATU.  
 DR InterPro; IPR002016; Peroxidase.  
 DR Pfam; PF00141; peroxidase.1.  
 DR PRINTS; PR00458; PEROXIDASE.  
 DR PROSITE; PS00436; PEROXIDASE\_2, 1.  
 FT NON\_TER  
 SQ SEQUENCE 97 AA; 10506 MW; A1862BE20C045E8E CRC64;

Query Match  
 Best Local Similarity 0.8%; Score 8; DB 10; Length 97;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 680 VTLAGPS 687  
 DB 82 VTLAGPS 89

## RESULT 5

OY 09PXT8 PRELIMINARY; PRT; 133 AA.

AC 09PXT8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE Genome polyprotein [contains: envelope glycoprotein E2/NS1 (GP68)]  
 DE (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NCBI\_TaxID=11103;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94069940; PubMed=8249288;  
 RA Higashi Y., Kakumu S., Yoshiooka K., Wakita T., Mizokami M., Ohba K.,  
 RA Ito Y., Ishikawa T., Takayanagi M., Nagai Y.;  
 RT "Dynamics of genome change in the E2/NS1 region of hepatitis C virus  
 RT in vivo."  
 RL Virology 197;659-668(1993).  
 DR InterPro; IPR002531; HCV NS1.  
 DR Pfam; PF01560; HCV NS1.  
 DR ProDom; PD186062; HCV NS1; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; Transmembrane.  
 SQ SEQUENCE 133 AA; 14623 MW; D04D3DC043550488 CRC64;

Query Match  
 Best Local Similarity 0.8%; Score 8; DB 12; Length 133;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 133 HGFASLSA 139  
 DB 26 HGFASLSA 33

## RESULT 6

OY 09CSR4 PRELIMINARY; PRT; 248 AA.

AC 09CSR4;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)

AT3949120.  
 GN AT3949120/T2J13\_40.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Shum P., Chen H., Cheuk R., Kim C.J., Banh J., Bowser L.,  
 RA Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y., Ishida J.,  
 RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B., Lee J.M.,  
 RA Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Pham P.K., Quach H.U., Sakano H., Sakurai T., Sacou M., Seki M.,  
 RA Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W.,  
 RA Theologis A., Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF324700; AAC40051.2; -  
 DR HSSP; P00433; 2ATU.  
 DR InterPro; IPR02016; Peroxidase.  
 DR Pfam; PF00141; peroxidase; 1.  
 DR PRINTS; PR00458; PEROXIDASE.  
 DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
 SQ SEQUENCE 248 AA; 27154 MW; 3007F7117D236BEA CRC64;

Query Match 0.8%; Score 8; DB 10; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 680 VTLAGSPS 687  
 Db 139 VTLAGSPS 146

RESULT 7  
 ID 08V7G7 PRELIMINARY; PRT; 302 AA.  
 AC 08V7G7;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE ORF3.  
 OS TT virus.  
 OS Viruses; ssDNA viruses; unclassified ssDNA viruses.  
 NCBI\_TaxID=68887;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=UT41F;  
 RC MEDLINE=21844401; PubMed=11855633;  
 RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,  
 RA Okamoto H.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=UT41F;  
 RC MEDLINE=21844401; PubMed=11855633;  
 RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,  
 RA Okamoto H.;  
 RT "Analysis of the complete genomes of thirteen TT virus variants  
 RT classified into the fourth and fifth genetic groups, isolated from  
 RT viremic infants.";  
 RL Arch. Virol. 147:21-41(2002).  
 DR EMBL; AB064603; BAB79340.1;  
 DR InterPro; IPR004118; TT\_ORF2.  
 DR Pfam; PF02957; TT\_ORF2; 1.  
 SQ SEQUENCE 302 AA; 32388 MW; CD492A959881A33C CRC64;

Query Match 0.8%; Score 8; DB 12; Length 302;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 DSAESTG 152  
 Db 278 DSAESTG 285

RESULT 8  
 ID Q42584 PRELIMINARY; PRT; 353 AA.  
 AC Q42584;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Peroxidase (EC 1.11.1.7).  
 GN PRXB3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CV. COLUMBIA;  
 RC MEDLINE=94159792; PubMed=8115548;  
 RA Intapruk C., Takano M., Shinozaki A.;  
 RT "Nucleotide sequence of a new cDNA for peroxidase from Arabidopsis  
 RT thaliana.";  
 RL Plant Physiol. 104:285-286(1994).  
 DR EMBL; X71794; CAA50677.1; -  
 DR HSSP; P00433; 2ATU.  
 DR InterPro; IPR02016; Peroxidase.  
 DR Pfam; PF00141; peroxidase; 1.  
 DR PRINTS; PR00458; PEROXIDASE.  
 DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
 KW Oxidoreductase; Peroxidase.  
 SQ SEQUENCE 353 AA; 38951 MW; 8ADC65C72E19810B CRC64;

Query Match 0.8%; Score 8; DB 10; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 680 VTLAGSPS 687  
 Db 139 VTLAGSPS 146

RESULT 9  
 ID Q9SMU8 PRELIMINARY; PRT; 353 AA.  
 AC Q9SMU8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Peroxidase (Putative peroxidase protein) (AT3949120/T2J13\_40).  
 GN T2J13\_40.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,  
 RA Wewes H.W., Lemcke K., Mayer K.F.X., Queitner F., Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Liu S.X., Sakano H., Pham P.K., Yamada K., Banh J., Egu P., Lee J.M.,  
 RA Toriumi M., Yu G., Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene T2J13\_40 (GI:6522554).";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]

RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
 RA Carinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shimizu P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Full length cDNA of gene T2J13.40 (GI:5522554)."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carinci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.,  
 RT "Arabidopsis cDNA clones."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Meyers M.C., Banh J., Bowser L.,  
 RA Carinci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Mu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.,  
 RT "Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases."  
 RL EMBL: AL132967; CAB61998.1; -;  
 DR EMBL: AF326880; AAC4162.1; -;  
 DR EMBL: AF329700; AKK0382.1; -;  
 DR EMBL: AF419569; AAL31901.1; -;  
 DR EMBL: AY079106; AAL8490.1; -;  
 DR HSSP: P00433; 2NTU.  
 DR InterPro: IPR002016; Peroxidase.  
 DR Pfam: PF00141; peroxidase; 1.  
 DR PRINTS: PR00458; PEROXIDASE.  
 DR PROSITE: PS00435; PEROXIDASE 1; 1.  
 DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
 DR Peroxidase.  
 SQ SEQUENCE 353 AA; 38832 MW; 8CDB98A2EF3E130B CRC64;  
 Query Match 0.8%; Score 8; DB 10; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 680 VTIAGSPS 687  
 DB 139 VTIAGSPS 146  
 RESULT 10  
 ID 09A0D7 PRELIMINARY; PRT; 430 AA.  
 AC 09A0D7;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical protein SPY0818.  
 GN SPY0818.  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 OX NCB1\_Taxid=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;  
 RX MEDLINE=21192684; PubMed=11296296;

RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Primeaux C., Sezter S., Svorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.,  
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).  
 DR EMBL: AE006532; AKK33754.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 430 AA; 46538 MW; 5BB9B2223C45B436 CRC64;  
 Query Match 0.8%; Score 8; DB 16; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 991 RGLMDLIT 998  
 DB 344 RGLMDLIT 351  
 RESULT 11  
 ID 0921I4 PRELIMINARY; PRT; 444 AA.  
 AC 0921I4;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE TOLB protein.  
 GN TOLB OR RC0406.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCB1\_Taxid=781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=MALISH 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audifren P., Fournier P.-E., Barbe V.,  
 RA Samsen D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.,  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."  
 RL Science 293:2093-2098 (2001).  
 DR EMBL: AE008605; AAL2944.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 444 AA; 49080 MW; 3A0EE0543B53D7E2 CRC64;  
 Query Match 0.8%; Score 8; DB 16; Length 444;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 161 PRGDYIAP 166  
 DB 341 PRGDYIAP 348  
 RESULT 12  
 ID 092891 PRELIMINARY; PRT; 695 AA.  
 AC 092891;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein CPN0458.  
 GN CPN0458.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCB1\_Taxid=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lamell C., Fan J., Hyman R.W.,  
 RA Ollinger L., Grimwood J., Davis R.W., Stephens R.S.,  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";

RL Nat. Genet. 21:385-389 (1999).  
 DR EMBL: A8001630; AAD18600.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 695 AA; 78562 MW; 446CFEB1014693B CRC64;  
 Query Match 0.8%; Score 8; DB 16; Length 695;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 868 VSSCVAGI 875  
 DB 67 VSSCVAGI 74  
 RESULT 13  
 09J558 PRELIMINARY; PRT; 695 AA.  
 AC 09J558.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE Hypothetical protein CP00458.  
 GN CP00458 OR CP0294.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NC NCB1\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,  
 Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 Winn M., Nelson W., Deboy R., Kolony J., McClarty G., Salzberg S.L.,  
 Eisen J., Fraser C.M.;  
 RA "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 pneumoniae AR39."  
 RT Nucleic Acids Res. 28:1397-1406 (2000).  
 RL Nucleic Acids Res. 28:1397-1406 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=U138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWR029 from USA."  
 RT Nucleic Acids Res. 28:2311-2314 (2000).  
 RL EMBL: A602191; AAF38151.1; -.  
 DR EMBL: A602191; AAF38151.1; -.  
 DR TIGR: CP0294; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 695 AA; 78471 MW; 1F9625881081B670 CRC64;  
 Query Match 0.8%; Score 8; DB 16; Length 695;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 868 VSSCVAGI 875  
 DB 67 VSSCVAGI 74  
 RESULT 14  
 042958 PRELIMINARY; PRT; 744 AA.  
 AC 042958.  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DE Hypothetical 82.9 kDa protein C19G7.10C in chromosome II.  
 GN SPB1967.10C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetes; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NC NCB1\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Lyne M., Rajandream M.A., Barrell B.G., Devlin K., Churcher C.M.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1 SIMILARITY: TO YEAST PAT1.  
 DR EMBL: AL021839; CAA17064.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 744 AA; 82920 MW; 888E06D842FD92B5 CRC64;  
 Query Match 0.8%; Score 8; DB 3; Length 744;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 183 LKQSTVN 190  
 DB 642 LKQSTVN 649  
 RESULT 15  
 095XN6 PRELIMINARY; PRT; 77 AA.  
 AC 095XN6.  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DE Hypothetical 8.6 kDa protein.  
 GN Y71G12B.14.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Rhabditinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium."  
 RL Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Du H., Lamar B., Wohlmann P.;  
 RA "The sequence of C. elegans cosmid Y71G12B."  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RA "Direct Submission."  
 RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AC025726; AAK73913.1; -.  
 DR InterPro: IPR001163; snRNP\_Sm.  
 DR Pfam: PF01423; Sm; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 77 AA; 8602 MW; 805A43965E96C978 CRC64;  
 Query Match 0.7%; Score 7; DB 5; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 237 RGNVLY 243  
 DB 65 RGNVLY 71  
 RESULT 16  
 09W2P5 PRELIMINARY; PRT; 79 AA.  
 ID 09W2P5

AC Q9WBP5; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CG9344 protein.  
 GN CG9344.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaralides P.G., Scherer S.E., Li P.W., Hopkins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Champs M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bower B.P., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA Deason K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Fodor K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
 RA Jatala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlechina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Follard J., Puri V., Reese M.G.,  
 RA Reiner K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195 (2000).  
 DR EMBL:AE003452; AAF64645.1; -  
 DR FlyBase: FBgn0034564; CG9344.  
 DR InterPro: IPR001163; snRNP\_Sm.  
 DR Pfam: PF01423; Sm; 1.  
 SQ SEQUENCE 79 AA; 9023 MW; 0480FAC624ADAF5B CRC64;

Query Match 0.7%; Score 7; DB 5; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 RGNNTVLY 243  
 DB 65 RGNNTVLY 71

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Putative small nuclear ribonucleoprotein polypeptide F.  
 GN ATG243810.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosida II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Baul C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanden K., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrara A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copeland G.P., Preuss D., Nieman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana."  
 RL Nature 402:761-766 (1999).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA;  
 RX Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC002333; AAB64025.1; -  
 DR InterPro: IPR001163; snRNP\_Sm.  
 DR Pfam: PF01423; Sm; 1.  
 KW Nucleocapsid; Ribonucleoprotein.  
 SQ SEQUENCE 91 AA; 9728 MW; 22DFB2C20E23B8 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 RGNNTVLY 243  
 DB 73 RGNNTVLY 79

RESULT 18  
 ID Q9M123 PRELIMINARY; PRT; 91 AA.  
 AC Q9M123;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE U6 snRNA-associated Sm-like protein.  
 GN P24316.80.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosida II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA;  
 RX Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL138647; CAB75800.1; -  
 DR InterPro: IPR001163; snRNP\_Sm.  
 DR Pfam: PF01423; Sm; 1.  
 SQ SEQUENCE 91 AA; 9913 MW; 950E3FE402F92045 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 63;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 RGNVLY 243  
Db 73 RGNVLY 79

## RESULT 19

Q9F1F2 PRELIMINARY; PRT; 93 AA.  
AC Q9F1F2; 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)  
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)  
DN Hypothetical 10.4 KDa protein.  
GN EP0044.  
OS Enterococcus faecalis (Streptococcus faecalis).  
OC Plasmid pAM373.  
CC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Enterococcaceae; Enterococcus.  
OX NCBI\_TaxID=1351;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20453452; PubMed=10998166;  
RA De Boever E.H., Clewell D.B., Fraser C.M.;  
RT "Enterococcus faecalis conjugative plasmid pAM373: complete nucleotide  
sequence and genetic analyses of sex pheromone response."  
RL Mol. Microbiol. 37:1327-1341(2000).  
DR EMBL; AE002565; AAG40455.1;  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 93 AA; 10366 MW; 344C9A25E2DD1922 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 YADKGS 289  
Db 50 YADKGS 56

RESULT 20  
Q9XJPI PRELIMINARY; PRT; 100 AA.

AC Q9XJPI; 01-NOV-1999 (TREMblrel. 12, Created)  
DT 01-NOV-1999 (TREMblrel. 12, last sequence update)  
DE 01-JUN-2002 (TREMblrel. 21, last annotation update)  
DN Proliferating cell nuclear antigen (Fragment).  
GN PCNA.  
OS Aureococcus anophagefferens.  
OC Eukaryota; stramenopiles; Pelagophyceae; Aureococcus.  
OX NCBI\_TaxID=44056;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PT-1;  
RA Lin S.;  
RT "Proliferating cell nuclear antigen in the brown tide alga."  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF126735; AAD29399.1;  
DR HSBP; P12004; IAXC.  
DR InterPro; IPR000730; Pr\_cel\_nuc\_antig.  
DR Pfam; PF02747; PCNA\_C; 1.  
DR ProDom; PD002673; Pr\_cel\_nuc\_antig; 1.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 100 AA; 11104 MW; 0579A7DA2EF0C773 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 100;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 ESLGIPD 794

Db 13 ESLGIPD 19

## RESULT 21

Q97RY2 PRELIMINARY; PRT; 106 AA.  
AC Q97RY2; 01-OCT-2001 (TREMblrel. 18, Created)  
DT 01-OCT-2001 (TREMblrel. 18, last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DN Hypothetical protein SP0650.  
GN SP0650.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TR64;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tetelini H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., Deboy R.T., Holt D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtapple E., Khouri H., Wolf A.M., Utecht T.R., Hansen C.B.,  
RA McDonald L.A., Feldblyum T.V., Angiolini S., Dickinson T., Hickey E.K.,  
RA Holt I.B., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
pneumoniae."  
RL Science 293:498-506(2001).  
DR EMBL; AE007374; AAK74796.1;  
DR TIGR; SP0650;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 106 AA; 12657 MW; 4C3288EB5A4B572 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 106;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 MILTLV 476  
Db 1 MILTLV 7

RESULT 22  
Q9N7S4 PRELIMINARY; PRT; 115 AA.

AC Q9N7S4; 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)  
DE 01-OCT-2000 (TREMblrel. 15, last annotation update)  
DN Possible kila0568 protein (Fragment).  
GN LM28.111.  
OS Leishmania major.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIDELIN;  
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL390935; CAC00863.1;  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 115 AA; 13150 MW; 7D3510574146C72D CRC64;

Query Match 0.7%; Score 7; DB 5; Length 115;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 LADRLIG 768



Db 62 LADRLIG 68

## RESULT 23

Q9M280 PRELIMINARY; PRT; 116 AA.  
AC Q9M280; 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
GN Hypothetical 12.7 kDa protein.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W., Rudd S.,  
RA Lemcke K., Mayer K.F.X., Quettler F., Salanoubat M.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL138641; CAB86923.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 116 AA; 12696 MW; 6E43AAD0115740F6 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 116;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 434 ETTVLSG 440  
DB 34 ETTVLSG 40

## RESULT 24

Q9RAN1 PRELIMINARY; PRT; 122 AA.  
AC Q9RAN1; 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Membrane protein.  
GN ORF-1.  
OS Methylobacillus flagellatum.  
OC Bacteria; Proteobacteria; beta subdivision; Methylobacillus group;  
OC Methylobacillus.  
OX NCBI\_TaxID=405;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=KT;  
RX MEDLINE=20055654; PubMed=10589737;  
RA Marchenko G.N., Marchenko N.D., Tsygankov Y.D., Chistoseedov A.Y.;  
RT "Organization of theonine biosynthesis genes from the obligate  
RL methylobiotroph Methylobacillus flagellatus.";  
DR EMBL; L78665; AAF21127.1; -  
SQ SEQUENCE 122 AA; 13500 MW; 7D8E0F4621303693 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 122;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 827 PGSLILP 833  
DB 37 PGSLILP 43

## RESULT 25

Q94045 PRELIMINARY; PRT; 124 AA.  
AC Q94045; 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
GN T13F2.2 protein.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Swindburne J.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology.";  
RL Science 282:2012-2018 (1998).  
DR EMBL; Z81122; CAB03353.1; -  
DR HSSP; PS3999; 1PCF.  
DR InterPro; IPR003173; PC4.  
DR Pfam; PF02229; PC4; 1.  
SQ SEQUENCE 124 AA; 14446 MW; 1B58475E1A659D91 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 124;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 609 YYIDRDS 615  
DB 87 YYIDRDS 93

## RESULT 26

Q96404 PRELIMINARY; PRT; 132 AA.  
AC Q96404; 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE Erythrocyte membrane protein 1 (Fragment).  
GN PCR351.2VAR8.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=56833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chen Q., Wahlgren M.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF039280; AAD02163.1; -  
FT NON\_TER 1 132  
FT NON\_TER 1 132  
SQ SEQUENCE 132 AA; 15584 MW; 2532ADBD04E1B63 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 132;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 529 KGSKGKQ 535  
DB 85 KGSKGKQ 91

## RESULT 27

Q97UE3 PRELIMINARY; PRT; 134 AA.  
AC Q97UE3;

DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE Hypothetical protein SSO3078.  
 GN SSO3078.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 CC Sulfolobus.  
 NCBI\_TaxID=2287;  
 RN NCBI\_TaxID=2287;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RA MEDLINE=2133296; PubMed=1142726;  
 RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aways M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moers A., Erasmo G., Fletcher C., Gordon P.M.K.,  
 RA Helkamp-de Jong J., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL; AE006899; AK43178.1; -.  
 DR InterPro; IPR002716; PIN.  
 DR Pfam; PF01850; PIN; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 134 AA; 15459 MW; E0CAC16D74E039A5 CRC64;

Query Match 0.7%; Score 7; DB 17; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 89;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 714 CTDNVD 720  
 Db 76 CTDNVD 82

RESULT 28  
 Q9EUI6 PRELIMINARY; PRT; 138 AA.  
 AC Q9EUI6;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Orf40.  
 OS Salmonella enterica subsp. enterica serovar Choleraesuis.  
 OC Plasmid 50k virulence plasmid.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxID=119912;  
 RN NCBI\_TaxID=119912;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RE-1;  
 RA MEDLINE=21153631; PubMed=11254626;  
 RA Haneida T., Okada N., Nakazawa N., Kawakami T., Danbara H.;  
 RT "Complete DNA Sequence and Comparative Analysis of the 50-Kilobase  
 RT Virulence Plasmid of Salmonella enterica Serovar Choleraesuis.";  
 RL Infect. Immun. 69:2612-2620(2001).  
 DR EMBL; AB040415; BAB20547.1; -.  
 KW Plasmid.  
 SQ SEQUENCE 138 AA; 16387 MW; C1E38C2757422F62 CRC64;  
 Qy Query Match 0.7%; Score 7; DB 2; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 93 NKEVARI 99

RESULT 29  
 005567

ID 005567 PRELIMINARY; PRT; 141 AA.  
 AC 005567;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE Hypothetical 15.8 kDa protein.  
 GN MLC33.17.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1769;  
 RN NCBI\_TaxID=1769;  
 RP SEQUENCE FROM N.A.  
 RA Badcock K., Churche C.M.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Barrett B.G., Rajadream M.A.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9318700; PubMed=8446027;  
 RA Eiglmeyer K., Honore N., Woods S.A., Caudron B., Cole S.T.;  
 RT "Use of an ordered cosmid library to deduce the genomic organization  
 RT of Mycobacterium leprae.";  
 RL Mol. Microbiol. 7:197-206(1993).  
 DR EMBL; Z94723; CAB08133.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 141 AA; 15792 MW; 0430A69020B7988 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 229 EFHSYEL 235  
 Db 83 EFHSYEL 89

RESULT 30  
 Q9HRJ4 PRELIMINARY; PRT; 159 AA.  
 AC Q9HRJ4;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE Vng0670h.  
 OS Halobacterium sp. (strain NRC-1).  
 GN Vng0670h.  
 OC Archaea; Euryarchaeota; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 NCBI\_TaxID=64091;  
 RN NCBI\_TaxID=64091;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Laasy S.R., Baliga N.S., Thorsen V., Shroga J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL; AB005013; AAG19164.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 159 AA; 15971 MW; BCC586A96222C254 CRC64;

Query Match 0.7%; Score 7; DB 17; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 28 LLLMAGT 34  
 Db 124 LLLMAGT 130

## RESULT 31

Oy 090732 PRELIMINARY; PRT; 163 AA.  
 ID 090732;  
 AC 090732;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Oocyst wall protein (Fragment).  
 OS Cryptosporidium serpentis.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
 OC Cryptosporidiidae; Cryptosporidium.  
 NCBI\_TaxID=83399;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99437830; PubMed=10508071;  
 RA Sulaiman I.M., Xiao L., Lal A.A.;  
 RT "Evaluation of cryptosporidium parvum genotyping techniques."; Appl. Environ. Microbiol. 65:4431-4435(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Sulaiman I.M., Xiao L., Lal A.A.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF161580; AAF04006.2; -  
 DR InterPro; IPR002086; Aldehyde dehydr.  
 DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 163  
 SQ SEQUENCE 163 AA; 17221 MW; SDDAE831D52FPA CRC64;

Query Match 0.7%; Score 7; DB 5; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 32

Oy 090732 PRELIMINARY; PRT; 163 AA.  
 ID 090732;  
 AC 090732;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Oocyst wall protein (Fragment).  
 OS Cryptosporidium muris.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
 OC Cryptosporidiidae; Cryptosporidium.  
 NCBI\_TaxID=5808;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20551138; PubMed=11097936;  
 RA Xiao L., Limor J., Morgan U.M., Sulaiman I.M., Thompson R.C.A., Lal A.A.;  
 RT "Sequence Differences in the diagnostic Target Region of the Oocyst Wall Protein Gene of Cryptosporidium Parasites."; Appl. Environ. Microbiol. 66:5499-5502(2000).  
 DR EMBL; AF266264; AAC39045.1; -  
 DR InterPro; IPR002086; Aldehyde dehydr.  
 DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 163  
 SQ SEQUENCE 163 AA; 17295 MW; 734B427DCFB9D19E CRC64;

Query Match

0.7%; Score 7; DB 5; Length 163;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 621 CPNTIL 627  
 Db 96 CPNTIL 102

## RESULT 33

Oy 083448 PRELIMINARY; PRT; 165 AA.  
 ID 083448;  
 AC 083448;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Disease resistance-like protein 585-6 (Fragment).  
 OS Mentha longifolia (Horse mint).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Lamiales; Lamiales; Mentha.  
 NCBI\_TaxID=38859;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Davis T.M., Vining K.J., Smith C.A., Zhang Q.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF469686; AAL83215.1; -  
 FT NON\_TER 1  
 FT NON\_TER 165  
 SQ SEQUENCE 165 AA; 18180 MW; 61596CB7937AFDPC CRC64;

Query Match 0.7%; Score 7; DB 10; Length 165;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 34

Oy 0833X6 PRELIMINARY; PRT; 165 AA.  
 ID 0833X6;  
 AC 0833X6;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Disease resistance-like protein 17-36 (Fragment).  
 OS Mentha longifolia (Horse mint).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Lamiales; Lamiales; Mentha.  
 NCBI\_TaxID=38859;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Davis T.M., Vining K.J., Smith C.A., Zhang Q.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF474173; AAL83251.1; -  
 FT NON\_TER 1  
 FT NON\_TER 165  
 SQ SEQUENCE 165 AA; 18312 MW; 91B59DD313239C2 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 165;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 787 LESLIGIP 793  
 Db 87 LESLIGIP 93

## RESULT 35

Oy 090733 PRELIMINARY; PRT; 167 AA.  
 ID 090733

```

AC 09U733;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Oocyst wall protein (Fragment).
OS Cryptosporidium muris.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5808;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99437830; PubMed=10508071;
RA Sulaiman I.M., Xiao L., Lal A.A.;
RT "Evaluation of cryptosporidium parvum genotyping techniques.";
RL Appl. Environ. Microbiol. 65:4431-4435(1999).
DR EMBL; AF161579; AAF04005.1; -.
DR InterPro; IPR002086; Aldehyde dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 167
SQ SEQUENCE 167 AA; 17665 MW; DA5BAFD3434E427D CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 621 CPNTIL 627
DB 96 CPNTIL 102

RESULT 36
Q9GGE1 PRELIMINARY; PRT; 167 AA.
ID 09GGE1;
AC 09GGE1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Oocyst wall protein (Fragment).
OS OMP.
OC Cryptosporidium andersoni.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=117008;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=845;
RA Xiao L., Limor J., Morgan U.M., Sulaiman I.M., Thompson R.C.A.,
RA Lal A.A.;
RT "Sequence Differences in the Diagnostic Target Region of the Oocyst
RT Wall Protein Gene of Cryptosporidium Parasites.";
RL Appl. Environ. Microbiol. 66:5499-5502(2000).
DR EMBL; AF266262; AAG39043.1; -.
DR InterPro; IPR002086; Aldehyde dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 167
SQ SEQUENCE 167 AA; 17623 MW; 2E6AA3161B4243F7 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 621 CPNTIL 627
DB 97 CPNTIL 103

RESULT 37
Q9GGE1 PRELIMINARY; PRT; 167 AA.
ID 09GGE1;
AC 09GGE1;

```

```

DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Oocyst wall protein (Fragment).
OS OMP.
OC Cryptosporidium serpentis.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=83999;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=754;
RA Xiao L., Limor J., Morgan U.M., Sulaiman I.M., Thompson R.C.A.,
RA Lal A.A.;
RT "Sequence Differences in the Diagnostic Target Region of the Oocyst
RT Wall Protein Gene of Cryptosporidium Parasites.";
RL Appl. Environ. Microbiol. 66:5499-5502(2000).
DR EMBL; AF266275; AAG39056.1; -.
DR InterPro; IPR002086; Aldehyde dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 167
SQ SEQUENCE 167 AA; 17623 MW; 2E6AA3161B4243F7 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 621 CPNTIL 627
DB 97 CPNTIL 103

RESULT 38
Q9S709 PRELIMINARY; PRT; 173 AA.
ID 09S709;
AC 09S709;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE Hypothetical protein.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartiidae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 1, PAC
RT clone, p071810."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 1, PAC
RT clone, p0705D01."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP000570; BAA85199.1; -.
DR EMBL; AP000492; BAA84624.1; -.
SQ SEQUENCE 173 AA; 19923 MW; 636824223DDCC29D CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 535 QSYTYII 541
DB 131 QSYTYII 137

```

RESULT 39

ID Q9UTV4 PRELIMINARY; PRT; 188 AA.

AC Q9UTV4; 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DE Hypothetical 21.0 kDa protein (Fragment)

OS Schizosaccharomyces pombe (Fission yeast)

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC NCB1\_TaxID=4896;

RN NCB1\_TaxID=4896;

RP SEQUENCE FROM N.A.

RC STRAIN=968 H90;

RX MEDLINE=20223868; PubMed=10759889;

RA Ding D.Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,

RT "Large-scale screening of intracellular protein localization in living

RL fission yeast cells by the use of a GFP-fusion genomic DNA library."

DR EMBL; AB027973; BAA87277.1; -

DR HSSP; P21707; 1BYN

DR InterPro; IPR000008; C2.

DR Pfam; PF00168; C2; 1.

DR PRINTS; PR00360; C2DOMAIN.

DR SMART; SM00239; C2; 1.

DR PROSITE; PS00499; C2\_DOMAIN\_1; UNKNOWN\_1.

DR PROSITE; PS50004; C2\_DOMAIN\_2; 1.

KW Hypothetical protein.

FT NOY\_TER 1

FT NOY\_TER 1

SQ SEQUENCE 188 AA; 21013 MW; 59E63D8C2BED8750 CRC64;

Db 121 DLPADS 127

Query Match 0.7%; Score 7; DB 3; Length 188;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 954 DLPADS 960

Db 121 DLPADS 127

Query Match 0.7%; Score 7; DB 3; Length 188;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 954 DLPADS 960

Db 121 DLPADS 127

Query Match 0.7%; Score 7; DB 3; Length 188;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 954 DLPADS 960

Db 121 DLPADS 127

Query Match 0.7%; Score 7; DB 3; Length 188;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 954 DLPADS 960

Db 121 DLPADS 127

Query Match 0.7%; Score 7; DB 3; Length 188;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 41

ID Q9RWR8 PRELIMINARY; PRT; 197 AA.

AC Q9RWR8; 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DE Hypothetical protein DR0598.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.

OC NCB1\_TaxID=1299;

RN NCB1\_TaxID=1299;

RP SEQUENCE FROM N.A.

RC STRAIN=R1;

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,

RA Moffatt K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utecherback T., Zaleski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Frazer C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus

RL radiodurans R1."

DR EMBL; AE001918; AAF10178.1; -

DR TIGR; DR0598; -

DR InterPro; IPR002886; Peptidase M37.

DR Pfam; PF01551; Peptidase M37.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 197 AA; 21348 MW; 7E0A29BC0874CC50 CRC64;

Db 179 IYTAGA 185

Query Match 0.7%; Score 7; DB 16; Length 197;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 458 IYTAGA 464

Db 179 IYTAGA 185

Query Match 0.7%; Score 7; DB 16; Length 197;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 458 IYTAGA 464

Db 179 IYTAGA 185

Query Match 0.7%; Score 7; DB 16; Length 197;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 458 IYTAGA 464

Db 179 IYTAGA 185

Query Match 0.7%; Score 7; DB 16; Length 197;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 458 IYTAGA 464

Db 179 IYTAGA 185

Query Match 0.7%; Score 7; DB 16; Length 197;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 458 IYTAGA 464

Db 179 IYTAGA 185

Query Match 0.7%; Score 7; DB 16; Length 197;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DR PROSITE: PS00211; ABC TRANSPORTER; UNKNOWN 1.  
SQ SEQUENCE 210 AA; 21817 MW; 72C42BE3B1804D8A CRC64;

Query Match 0.7%; Score 7; DB 2; Length 210;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
412 .PAGEIPA 418  
195 .PAGEIPA 201

## RESULT 43

0931V4 PRELIMINARY; PRT; 211 AA.  
AC 0931V4;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Putative integral membrane protein.  
GN SC01316 OR SCBAC36F5.27C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Saunders D., Harris D.;  
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmid and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Beutley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)."  
RL Nature 417:141-147(2002).  
DR EMBL: AL592292; CAC42863.1; -  
DR InterPro: IPR000794; Ketoacyl-synt.  
DR Pfam: PF00597; Deda; 1.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; UNKNOWN 1.  
SQ SEQUENCE 211 AA; 22286 MW; 1A0F8EA70072A64A CRC64;

Query Match 0.7%; Score 7; DB 16; Length 211;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 787 LESTGIP 793  
db 25 LESTGIP 31

## RESULT 44

0821P0 PRELIMINARY; PRT; 215 AA.  
AC 0821P0;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Putative phosphoglycerate mutase (EC 5.4.2.1).  
GN GPMB OR YP00455.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;  
RX MEDLINE=21470413; PubMed=11586360.  
RA Parkhill J., Wren B.W., Thomson N.R., Titchall R.W., Holden M.T.G.,  
RA Prentice M.B., Septhalia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moulton S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
RL Nature 413:523-527(2001).  
DR EMBL: AJ414142; CAC89311.1; -  
DR InterPro: IPR001345; PG/BPGM\_mutase.  
DR Pfam: PF00300; PGAM; 1.  
DR PROSITE: PS00175; PG\_MUTASE; 1.  
DR K\* isomerase; Hypothetical protein; Complete proteome.  
SQ SEQUENCE 215 AA; 23829 MW; B59BC55444FA7FB CRC64;

Query Match 0.7%; Score 7; DB 16; Length 215;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 722 RIRPGES 728  
db 115 RIRPGES 121

## RESULT 45

09VV45 PRELIMINARY; PRT; 217 AA.  
AC 09VV45;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE CG12255 protein.  
GN CG12255  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoekstra R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abiliv J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunker B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
 RA Jallil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z.,  
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy C., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskens D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Palzer K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003527; AAF49477.1; -  
 DR FLYbase: FBgn0036618; CG12255.  
 DR InterPro: IPR000618; Insect cuticle.  
 DR Pfam: PF00379; Chitin bind 4; 1.  
 RT SEQUENCE 217 AA; 24162 MW; D362488E9F996A4 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 343 DANGETO 349  
 Db 74 DANGETO 80

RESULT 46  
 Q94LR6 PRELIMINARY; PRT; 223 AA.  
 AC 094LR6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE Hypothetical 23.5 kDa protein.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 NCBI\_Taxid=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPONBARE;  
 RA Buehl C.R., Yuan Q., Ouyang S., Mofatt K.S., Hill J.N., Gansberger K.,  
 RA Bremner M., Burgess S., Hance M., Shvartsbeyn M., Tselirin T.,  
 RA Riggs F., Hsiao J., Zismann V., Blunt S., Pai G., Vanken S.B.,  
 RA Uteback T.R., Feldblyum T.V., Quackenbush J., Salzberg S.L.,  
 RA White O., Frazer C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSUNB0011A08 genomic sequence.";  
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBD databases.  
 DR EMBL: AC034258; AAK54290.1; -  
 KW Hypothetical protein.  
 SO SEQUENCE 223 AA; 23534 MW; 02C547E08745A473 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 460 TAAASD 466  
 Db 35 TAAASD 41

RESULT 47  
 Q9ZK02 PRELIMINARY; PRT; 223 AA.  
 AC 09ZK02;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DE Putative unopispyrithogen-III synthase.  
 GN HMD OR UHP1145.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99)  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 NCBI\_Taxid=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
 RA Tummino P.J., Caruso A., Ulla-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Metberg D., Mills S.D., Ullang O., Taylor D.E., Vovis G.F.,  
 RA Trust T.J.;  
 RT "Genomic sequence comparison of two unrelated isolates of the human  
 RT gastric pathogen *Helicobacter pylori*.";  
 RL Nature 397:176-180(1999).  
 CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT  
 CC BINDS TWO MOLES OF GTP. ONE AT AN EXCHANGEABLE SITE ON THE BETA  
 CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.  
 DR EMBL: AE001542; AAD06731.1; -  
 DR InterPro: IPR002453; Beta\_tubulin.  
 DR Pfam: PF02602; HEM4; 1.  
 DR PROSITE: PS00228; TUBULIN\_B AUTOREG; 1.  
 KW GTP-binding; Microtubules; Complete proteome.  
 RT SEQUENCE 223 AA; 25325 MW; E857C9D8F676D9D CRC64;

Query Match 0.7%; Score 7; DB 16; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 973 LIFTSKN 979  
 Db 38 LIFTSKN 44

RESULT 48  
 Q9S0D9 PRELIMINARY; PRT; 224 AA.  
 AC 09S0D9;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DE Outer surface protein F.  
 GN BBR42.  
 OS *Borrelia burgdorferi* (Lyme disease spirochete).  
 OG Plasmid cp32-4.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; *Borrelia*.  
 NCBI\_Taxid=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Lathigra R.,  
 RA White O., Dodson R., Hickey E.K., Gwinn M., Peterson J., van Vugt R.,  
 RA Palmer N., Haft D., Rosa P., Stevenson B.;  
 RT "A bacterial genome in flux: The twelve linear and nine circular  
 RT extrachromosomal DNAs in an infectious isolate of the Lyme disease  
 RT spirochete *Borrelia burgdorferi*.";  
 RL Mol. Microbiol. 0:0-0(1999).

DR EMBL; AE001577; AAF07488.1; -  
 DR InterPro; IPR003483; OsPEF.  
 DR Pfam; PF02471; OsPEF; 1.  
 KW PfamId.  
 SQ SEQUENCE 224 AA; 25416 MW; 4C3AD3F10DC9CD11 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 363 DLEGAVK 369  
 Db 28 DLEGAVK 34

RESULT 49  
 ID 093803 PRELIMINARY; PRT; 224 AA.  
 AC 093803;  
 DT 01-DEC-2001 (TRENBLREL. 19, Created)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
 DE Putative regulatory protein.  
 GN SC06286 OR SCIG7.12C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycetaceae; Streptomyces.  
 OC NCBI\_TaxId=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K., Harris D.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA MEDLIN=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Knaash H., Hopwood D.A.;  
 RL "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabbittowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RL "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 CC -1- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 DR EMBL; AL591083; CAC37887.1; -  
 DR InterPro; IPR001647; HTH\_Tetr.  
 DR Pfam; PF00440; tetr.1.  
 KW DNA-binding; Transcription regulation.  
 SQ SEQUENCE 224 AA; 24083 MW; B341BD44A209C84 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 24 RLRLRL 30  
 Db 180 RLRLRL 186

RESULT 50  
 ID 042826 PRELIMINARY; PRT; 226 AA.  
 AC 042826;  
 DT 01-JUN-1998 (TRENBLREL. 06, Created)  
 DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)  
 DT 01-OCT-2001 (TRENBLREL. 18, Last annotation update)  
 DE pdh1 protein precursor.  
 GN PDH1 OR SPC1235.08C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OC NCBI\_TaxId=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JY336;  
 RX MEDLINE=98162726; PubMed=9501995;  
 RA Iba H., Takimoto M., Danjoh I., Fujiyama A.;  
 RT "Identification and characterization of a novel trans-membrane protein  
 gene, pdh1, from Schizosaccharomyces pombe.";  
 RL DNA Res. 4:393-396(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Wood V., Rajandream M.A., Barrell B.G., Murphy L., Harris D.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 DR EMBL; D88384; BAA2946.1; -  
 DR EMBL; AL031764; CAA2112.1; -  
 KW Transmembrane; Signal.  
 FT CHAIN 1 26 POTENTIAL.  
 FT SIGNAL 1 26 PDH1 PROTEIN.  
 FT DOMAIN 27 226 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 42 62 POTENTIAL.  
 FT DOMAIN 63 65 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 66 86 POTENTIAL.  
 FT DOMAIN 87 97 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 98 118 POTENTIAL.  
 FT DOMAIN 119 191 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 192 212 POTENTIAL.  
 FT DOMAIN 213 226 EXTRACELLULAR (POTENTIAL).  
 SQ SEQUENCE 226 AA; 26021 MW; 3C71DB50ECB317B1 CRC64;

Query Match 0.7%; Score 7; DB 3; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 783 ELFHLES 789  
 Db 214 ELFHLES 220

RESULT 51  
 ID 025822 PRELIMINARY; PRT; 226 AA.  
 AC 025822;  
 DT 01-JAN-1998 (TRENBLREL. 05, Created)  
 DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)  
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)  
 DE Uroporphyrinogen III COSYNTHASE (HEMD).  
 GN HP1224.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OC NCBI\_TaxId=210;  
 RN [1]



SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RA MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Mallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*  
 RT *pylori*.";  
 RL Nature 388:539-547(1997).  
 CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT  
 CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA  
 CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.  
 CC EMBL; AE000628; AAD08269.1; -.  
 DR TIGR; HP1224; -.  
 DR InterPro; IPR002453; Beta\_tubulin.  
 DR InterPro; IPR003754; HEMA4.  
 DR Pfam; PF02602; HEMA4; 1.  
 DR PROSITE; PS00228; TUBULIN\_B\_AUTOREG; 1.  
 DR GTP-binding; Hypothetical protein; Microtubules; Complete proteome.  
 SQ SEQUENCE 226 AA; 25862 MW; 82BAF4B54D6AFEBF CRC64;

Query Match 0.7%; Score 7; DB 16; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 973 LIFTSKN 979  
 Db 38 LIFTSKN 44

RESULT 52  
 Q9MP30 PRELIMINARY; PRT; 241 AA.  
 AC 09MP30;  
 DT 01-OCT-2000 (TRENBLREL. 15, Created)  
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE Cytochrome b (Fragment).  
 OS *Forulis chalybaeus*.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
 OC Formicidae; Dolichoderinae; Forulius.  
 NCBI\_TaxID=121505;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ACLP;  
 RA Chiotis M., Jermiin L.S., Crozier R.H.;  
 RT "A molecular framework for the phylogeny of the ant subfamily  
 RT Dolichoderinae.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).  
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
 CC EMBL; AF146718; AAF66718.1; -.  
 DR InterPro; IPR00179; Cyt b b6.  
 DR Pfam; PF00032; cytochrome b C; 1.  
 DR Pfam; PF00033; cytochrome\_b\_N; 1.

DR PROSITE; PS00193; CYTOCHROME B CO; UNKNOWN.1.  
 KM Electron transport; Heme; Mitochondrion; Respiratory chain;  
 KW Transmembrane.  
 FT NON TER  
 SQ SEQUENCE 241 AA; 28400 MW; 2105B5735635D3E2 CRC64;

Query Match 0.7%; Score 7; DB 8; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 469 FMILTLV 475  
 Db 46 FMILTLV 52

RESULT 53  
 Q8SUT3 PRELIMINARY; PRT; 246 AA.  
 AC 08SUT3;  
 DT 01-JUN-2002 (TRENBLREL. 21, Created)  
 DT 01-JUN-2002 (TRENBLREL. 21, Last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
 DE Hypothetical protein ECU08\_0300.  
 GN ECU08\_0300.  
 OS *Encephalitozoon cuniculi*.  
 OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.  
 NCBI\_TaxID=6035;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA Genoscope;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RX MEDLINE=21576510; PubMed=11719806;  
 RA Katinka M.D., Duprat S., Cornillot E., Meunier G., Thonart F.,  
 RA Prensler G., Barbe V., Peyretailade E., Brotier P., Wincker P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT *Encephalitozoon cuniculi*.";  
 RL Nature 414:450-453(2001).  
 DR EMBL; AL590448; CAD26335.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 246 AA; 28355 MW; 34C9A1D609E0A946 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 828 GSLILPG 834  
 Db 49 GSLILPG 55

RESULT 54  
 O07821 PRELIMINARY; PRT; 247 AA.  
 AC 007821;  
 DT 01-JUL-1997 (TRENBLREL. 04, Created)  
 DT 01-JUL-1997 (TRENBLREL. 04, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE Flavoprotein (Fragment).  
 OS *SOX*.  
 GN *Paracoccus denitrificans*.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Paracoccus.  
 NCBI\_TaxID=266;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB17;  
 RX MEDLINE=95014061; PubMed=7928987;

RA Modara C., Kostka S., Egert M., Kelly D.P., Friedrich C.;  
 RT "Identification and sequence analysis of soxB gene essential for  
 RT sulfur oxidation ability of Paracoccus denitrificans GB17.";  
 RL J. Bacteriol. 176:6188-6191 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB17;  
 RA Friedrich C.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB17;  
 RX MEDLINE=97405897; Pubmed=9260941;  
 RA Modara C., Bardschewsky F., Friedrich C.G.;  
 RT "Cloning and characterization of sulfite dehydrogenase, two c-type  
 RT cytochromes, and a flavoprotein of Paracoccus denitrificans GB17;  
 RT Essential role of of sulfite dehydrogenase in lithotrophic sulfur  
 RT oxidation.";  
 RL J. Bacteriol. 179:5014-5023 (1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB17;  
 RA Friedrich C.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB17;  
 RA Friedrich C.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X79242; CAA55826.1; -.  
 DR HSP; Q06530; IPCD.  
 FT NON TER 247 247  
 SQ SEQUENCE 247 AA; 26445 MW; 8F10A5B235A16621 CRC64;  
 Query Match 0.7%; Score 7; DB 2; Length 247;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 679 TVTLAGG 685  
 DB 114 TVTLAGG 120  
 RESULT 55  
 Q9N9P6 PRELIMINARY; PRT; 248 AA.  
 AC Q9N9P6;  
 DT 01-OCT-2000 (TRENBLREL. 15, Created)  
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)  
 DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)  
 DE Hypothetical 27.3 KDa protein.  
 GN L2802.04;  
 OS Leishmania major.  
 OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Aert R., Volckaert G., Ivens A.C., Quail M., Rajandream M.A.,  
 RA Barrell B.G.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RX MEDLINE=98146435; Pubmed=9477341;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 RA Smith D.F.;  
 RT "A physical map of the Leishmania major Friedlin genome.";  
 RL Genome Res. 8:135-145 (1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Aert R., Volckaert G., Ivens A.C., Quail M., Rajandream M.A.,

RA Barrell B.G.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL359774; CAB95238.2; -.  
 DR EMBL; AL359683; CAC37117.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 248 AA; 27281 MW; 86E72D0E70970576 CRC64;  
 Query Match 0.7%; Score 7; DB 5; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 398 PYGSYSN 404  
 DB 156 PYGSYSN 162  
 RESULT 56  
 Q9HFW6 PRELIMINARY; PRT; 258 AA.  
 AC Q9HFW6;  
 DT 01-MAR-2001 (TRENBLREL. 16, Created)  
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE Lipase (Fragment).  
 OS Penicillium expansum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lin L., Xie B.-F., Yang G.-Z., Shi Q.-Q., Xie L.-H., Wu S.-G.,  
 RA Wu X.-F.;  
 RT "Cloning and expression of Penicillium expansum lipase gene.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF284064; AAG22769.1; -.  
 DR InterPro; IPR002921; Lipase\_3.  
 DR Pfam; PF01764; Lipase\_3; 1.  
 FT NON TER 1  
 SQ SEQUENCE 258 AA; 27262 MW; 2953E19E431FF9A9 CRC64;  
 Query Match 0.7%; Score 7; DB 3; Length 258;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 236 NRGNNVL 242  
 DB 181 NRGNNVL 187  
 RESULT 57  
 Q8ZXK9 PRELIMINARY; PRT; 267 AA.  
 AC Q8ZXK9;  
 DT 01-MAR-2002 (TRENBLREL. 20, Created)  
 DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
 DE Hypothetical protein PAE1369.  
 GN PAE1369.  
 OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 OX NCBI\_TaxID=13773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX Pubmed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller U.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989 (2002).  
 DR EMBL; AE009818; AAL63440.1; -.  
 DR InterPro; IPR003794; DUF191.

DR InterPro; IPR002029; Peptidase\_S8.  
 DR Pfam; PF02642; DUF191.1.  
 DR PROSITE; PS00136; SUBTILASE ASP; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 267 AA; 29602 MW; 08362C2A1FBC293 CRC64;

Query Match 0.7%; Score 7; DB 17; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 GAVKLP 372  
 DB 22 GAVKLP 28

RESULT 58  
 ID 08STF1 PRELIMINARY; PRT; 268 AA.  
 AC 08STF1;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DE Hypothetical protein EC01\_0090 (Hypothetical protein EC01\_1640).  
 GN EC01\_0090 OR EC005\_1640.  
 OS Encephalitozoon cuniculi.  
 OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.  
 OX NCBI\_Taxid=6035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA Genoscope;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1.  
 RA MEDLINE=21576510; PubMed=11719806;  
 RX Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
 RA Prensier G., Barde V., Peyretallade E., Brotier P., Wincker P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissenbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT Encephalitozoon cuniculi.";  
 RL Nature 414:450-453 (2001).  
 DR EMBL; AL590450; CAD25919.1; -  
 DR EMBL; AL590445; CAD26684.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 268 AA; 30530 MW; 4BA693E33A4567D8 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CKESRY 55  
 DB 176 CKESRY 182

RESULT 59  
 ID 092T29 PRELIMINARY; PRT; 275 AA.  
 AC 092T29;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Purative modification methylase protein.  
 GN HEMK2 OR RB1344 OR SM21514.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSymb (megaplasmid 2).  
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_Taxid=382;  
 DE [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396508; PubMed=11481431;  
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,  
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,  
 RA Golding B., Puhler A.;  
 RT "The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-  
 RT fixing endosymbiont Sinorhizobium meliloti.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).  
 DR EMBL; AL603646; CAC49744.1; -  
 DR InterPro; IPR004556; Hemk.  
 DR InterPro; IPR002052; N6\_Mtase.  
 DR InterPro; IPR004838; NHCtransf.1.  
 DR InterPro; IPR000051; SAM\_bind.  
 DR TIGRfams; TIGR00536; hemk\_fam.1.  
 DR PROSITE; PS00105; AA\_TRANSF\_CLASS\_1; UNKNOWN\_1.  
 DR PROSITE; PS00092; N6\_MTASE; UNKNOWN\_1.  
 KW Methyltransferase; Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 275 AA; 29305 MW; EBC3460BEFAD114 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 275;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 EDLEGAV 368  
 DB 154 EDLEGAV 160

RESULT 60  
 ID 08YSU3 PRELIMINARY; PRT; 282 AA.  
 AC 08YSU3;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein A112989.  
 GN A112989.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_Taxid=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213 (2001).  
 DR EMBL; AP003591; BAB74688.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 282 AA; 31531 MW; 7BEC9ADCD7A119AB CRC64;

Query Match 0.7%; Score 7; DB 16; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 KAGVSSQ 758  
 DB 204 KAGVSSQ 210

RESULT 61  
 ID 08U7C6 PRELIMINARY; PRT; 284 AA.  
 AC 08U7C6;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE ABC transporter, membrane spanning protein.

GN ATU4523 OR AGR L 692.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OC NCBI\_TaxID=176299;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Seubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gilliet M., Grant C.,  
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58.";  
 RL Science 294:2317-2323(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194; Miller N., Blanchard M.,  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Gurolo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Houmlel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,  
 RA Woliam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
 RA Plangan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).  
 DR EMBL: AE009380; AAL45317.1; ALT\_INIT.  
 DR EMBL: AE008234; AAK89920.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 284 AA; 30723 MW; B5B8BFA6B25E6 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 284;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 920 ALLTVL 926  
 DB 257 ALLTVL 263

RESULT 62

Q9P451 PRELIMINARY; PRT; 285 AA.

ID Q9P451  
 AC Q9P451;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Alkaline lipase (EC 3.1.1.3) (Triacylglycerol lipase precursor).  
 OS Penicillium cyclopium, and  
 OS Penicillium expansum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
 OC NCBI\_TaxID=60167, 27334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P.cyclopium, STRAIN=PG37;  
 RA Wu M., Qian Z., Min T., Sun C., Huang W.;  
 RT "Cloning and sequencing of alkaline lipase cDNA from Penicillium  
 RT cyclopium PG37.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P.expansum, STRAIN=MMC20718;  
 RA Wu M., Qian Z., Min T., Sun C., Huang W.;  
 RT "Cloning and sequencing of alkaline lipase DNA from Penicillium  
 RT expansum MMC20718.";

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P.expansum, STRAIN=PF898;  
 RA Lin L., Xie B.F., Shi Q.Q., Lin Q.Y., Yang G.Z., Wu X.F., Xie L.H.,  
 RA Wu S.G.;  
 RT "Cloning and sequencing of triacylglycerol lipase from Penicillium  
 RT expansum PF898.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF274320; AAF82375.1; -;  
 DR EMBL: AF286685; AAF99329.1; -;  
 DR EMBL: AF330635; AAK07480.1; -;  
 DR InterPro: IPR002921; Lipase\_3.  
 DR Pfam: PF01764; Lipase\_3; 1.  
 KW Hydrolyase; Signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 285  
 FT SEQUENCE 285 AA; 30224 MW; C2D27B91C7557D14 CRC64;

Query Match 0.7%; Score 7; DB 3; Length 285;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 NRGNNVL 242  
 DB 208 NRGNNVL 214

RESULT 63

Q94295 PRELIMINARY; PRT; 294 AA.

ID Q94295  
 AC Q94295;  
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical 34.3 kDa protein.  
 DE T01C4.3.  
 GN T01C4.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Graves T., Wohlmann P.;  
 RT "The sequence of C. elegans cosmid T01C4.";  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U70858; AAB09178.1; -;  
 DR InterPro: IPR002767; GPCR\_Rhodopsin.  
 DR PROSITE: PS50262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 294 AA; 34252 MW; 37760F176CC82405 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 294;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 FMILTVL 475

DB 39 FMILLTV 45

RESULT 64

ID Q9AB93 PRELIMINARY; PRT; 294 AA.

AC Q9AB93;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Succinyl-CoA synthetase, alpha subunit.

GN CC0338.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI\_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CB15;

RX MEDLINE=21173698; PubMed=11259647;

RA Nielsen W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eissen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

RA Deboy R.I., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,

RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Usterback T., Tran K., Wolf A., Vamathevan J., Ermolova M., White O.,

RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.,

RA "Complete genome sequence of Caulobacter crescentus.",

RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

RL -1 SIMILARITY: TO OTHER ALPHA SUBUNITS OF SUCCINYL-COA SYNTHETASE, OF

CC MALATE-COA LIGASE AND TO ATP CITRATE-LYASE.

CC EMBL; AB005707; AAK2325.1; -.

DR HSPD; P07459; 1SCU.

DR TIGR; CC0338; -.

DR InterPro: IPR003781; COA binding.

DR InterPro: IPR003033; COA\_ligase.

DR Pfam; PF02629; COA\_binding; 1.

DR TIGRfam; TIGR01019; succoalpa; 1.

DR PROSITE; PS01216; SUCCINYL\_COA\_LIG\_1; 1.

DR ATP-binding; ligase; phosphorylation; tricarboxylic acid cycle;

KW Complete proteome.

SQ SEQUENCE 294 AA; 30138 MW; 2291AD957C99336A CRC64;

Query Match

Best Local Similarity 100.0%; Score 7; DB 16; Length 294;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 DPVKGT 90

DB 181 DPVKGT 187

RESULT 65

ID Q9K1F0 PRELIMINARY; PRT; 297 AA.

AC Q9K1F0;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Ffbr1.

GN Ffbr1.

OS Streptomyces hygroscopicus var. ascomyceticus.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=132248;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20323220; PubMed=10633099;

RA Wu K., Chung L., Revill W.P., Katz L., Reeves C.D.;

RT "The FK520 gene cluster of streptomyces hygroscopicus var.

RT ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual

polyketide extender units.";

RL Gene 251:81-90(2000).

CC -1 SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.

DR EMBL; AF235504; AAF6383.1; -.

DR InterPro: IPR000847; HTH\_LySR.

DR InterPro: IPR005119; LySR\_subst.

DR Pfam; PF00126; HTH\_1; 1.

DR Pfam; PF03466; LySR\_substrate; 1.

DR PRINTS; PR00039; HTH\_LYSR.

DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; UNKNOWN\_1.

KW DNA-binding; Transcription regulation.

SQ SEQUENCE 297 AA; 32462 MW; D015FBD4C59A19D CRC64;

Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 297;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 680 VTLGGP 686

DB 270 VTLGGP 276

RESULT 66

ID Q9BEC8 PRELIMINARY; PRT; 300 AA.

AC Q9BEC8;

DT 01-OCT-2001 (TrEMBLrel. 18, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Succinyl-CoA synthetase alpha-subunit.

GN M14303.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Phyllobacteriaceae; Mesorhizobium.

OX NCBI\_TaxID=381;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MAF303099;

RX MEDLINE=21082930; PubMed=11214968;

RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,

RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,

RA Takeuchi C., Yamada M., Tabata S.;

RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RT Mesorhizobium loti.",

RT DML Res. 7:331-338(2000).

RL EMBL; AB003003; BAB50992.1; -.

DR InterPro: IPR003781; COA binding.

DR InterPro: IPR003033; COA\_ligase.

DR Pfam; PF02629; COA\_binding; 1.

DR TIGRfam; TIGR01019; succoalpa; 1.

DR PROSITE; PS01216; SUCCINYL\_COA\_LIG\_1; 1.

DR PROSITE; PS00399; SUCCINYL\_COA\_LIG\_2; 1.

KW Complete proteome.

SQ SEQUENCE 300 AA; 30861 MW; 6FD24A5A474CE034 CRC64;

Query Match

Best Local Similarity 100.0%; Score 7; DB 16; Length 300;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 DPVKGT 90

DB 187 DPVKGT 193

RESULT 67

ID Q8ZLB0 PRELIMINARY; PRT; 300 AA.

AC Q8ZLB0;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

```

DE ABC superfamily (membrane), dipeptide transport protein 2.
GN DPPC OR STM3628.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
DR EMBL; AB008868; AAL22489.1; -.
DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF00528; BPD_transp; 1.
KW Complete proteome.
SQ SEQUENCE 300 AA; 3223 MW; 8856ECF3D4881F53 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 920 AILLTVL 926
DB 275 AILLTVL 281
|||||
Q8ZA21 PRELIMINARY; PRT; 300 AA.
ID Q8ZA21;
AC Q8ZA21;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Dipeptide transport system permease protein.
GN DPPC OR YPO4001.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Penrice M.B., Sebailha M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felkwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
DR EMBL; AJ141460; CAC93461.1; -.
DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF00528; BPD_transp; 1.
KW Complete proteome.
SQ SEQUENCE 300 AA; 32307 MW; F9F7F4C856E8EB13 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 920 AILLTVL 926
DB 275 AILLTVL 281
|||||

```

```

RESULT 69
Q8C61 PRELIMINARY; PRT; 300 AA.
ID Q8C61;
AC Q8C61;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE SucCyl1-SCA synthetase alpha chain.
GN SUCD OR ATU2637 OR AGR_C4779.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21608550; PubMed=11743193;
RX Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D., Sr.,
RA Chapman P., Clendening J., Decherge G., Gillet W., Grant C.,
RA Kutyavina T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).

Query Match 0.7%; Score 7; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 920 AILLTVL 926
DB 275 AILLTVL 281
|||||
Q8UC61 PRELIMINARY; PRT; 300 AA.
ID Q8UC61;
AC Q8UC61;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE SucCyl1-SCA synthetase alpha chain.
GN SUCD OR ATU2637 OR AGR_C4779.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21608550; PubMed=11743193;
RX Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D., Sr.,
RA Chapman P., Clendening J., Decherge G., Gillet W., Grant C.,
RA Kutyavina T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).

```

[2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Quinello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Hounell K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Garsen J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).  
 DR EMBL; AB009211; AAL43618.1; -  
 DR EMBL; AB008177; AAK8358.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 300 AA; 31126 MW; 03101D8C32CB3CA9 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 DPKGTE 90  
 DB 187 DPKGTE 193

## RESULT 71

092LJ5 PRELIMINARY; PRT; 300 AA.  
 AC 092LJ5.  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Probable succinyl-CoA synthetase alpha chain protein  
 DE (EC 6.2.1.5).  
 GN SUCD OR R03054 OR SMC02481.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Bolstead P., Becker A., Boutry M., Cadieu R., Dreano S., Gloux S.,  
 RA Godite T., Goffeau A., Kahn D., Kles B., Lelaure V., Masny D.,  
 RA Phil T., Portetelle D., Puhler A., Purnelle B., Ramepeyger U.,  
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gilbert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT Sinorhizobium meliloti strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 DR EMBL; AL591792; CAC47633.1; -  
 DR InterPro; IPR003781; CoA binding.  
 DR InterPro; IPR003781; CoA binding.  
 DR Pfam; PF02629; CoA binding; 1.  
 DR Pfam; PF00549; Ligase-CoA; 1.  
 DR TIGRPFAM; TIGR01019; succoAlpha; 1.  
 DR PROSITE; PS01216; SUCCINYL\_COA\_Lig\_1; UNKNOWN 1.  
 DR PROSITE; PS00399; SUCCINYL\_COA\_Lig\_2; UNKNOWN 1.  
 KW Ligase; Complete proteome.  
 SQ SEQUENCE 300 AA; 30927 MW; FB83DA110642AC20 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 DPKGTE 90  
 DB 187 DPKGTE 193

## RESULT 72

08YJES PRELIMINARY; PRT; 300 AA.  
 AC 08YJES.  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Succinyl-CoA synthetase alpha chain (EC 6.2.1.5).  
 GN BMEI0139.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;  
 RX MEDLINE=20020109; PubMed=11756688;  
 RA Delvecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,  
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
 RA Uablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,  
 RA Selkov E., Blzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
 RA Haselkorn R., Kyriades N., Overbeek R.;  
 RT "The genome sequence of the facultative intracellular pathogen  
 RT Brucella melitensis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 DR EMBL; AB009456; AAL51321.1; -  
 DR InterPro; IPR003781; CoA binding.  
 DR InterPro; IPR003781; CoA binding.  
 DR Pfam; PF02629; CoA binding; 1.  
 DR Pfam; PF00549; Ligase-CoA; 1.  
 DR TIGRPFAM; TIGR01019; succoAlpha; 1.  
 DR PROSITE; PS01216; SUCCINYL\_COA\_Lig\_1; 1.  
 DR PROSITE; PS00399; SUCCINYL\_COA\_Lig\_2; UNKNOWN 1.  
 KW Ligase; Complete proteome.  
 SQ SEQUENCE 300 AA; 31234 MW; 017C400CF7ADF81A CRC64;

Query Match 0.7%; Score 7; DB 16; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 DPKGTE 90  
 DB 187 DPKGTE 193

## RESULT 73

09Z799 PRELIMINARY; PRT; 303 AA.  
 AC 09Z799.  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE CT380 hypothetical protein.  
 GN CPN0807 OR CPJ0807.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shira M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).

DR EMBL; AE001662; AAD18945.1; -  
 DR EMBL; AP002548; BAA99015.1; -  
 DR InterPro; IPR000620; DUF6.  
 DR Pfam; PF00892; DUF6; 2.  
 KW Complete proteome.  
 SO SEQUENCE 303 AA; 33283 MW; A039571A51EB16B0 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 942 SKLVNNA 948  
 DB 20 SKLVNNA 26

## RESULT 74

ID O9KIS9 PRELIMINARY; PRT; 305 AA.

AC O9KIS9;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE Membrane protein, putative.  
 GN CP1064.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RC STRAIN=AR39;  
 RA MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey B.K., Peterson J., Ueberback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., Debey R., Koloney J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Frazer C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39."  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AE002263; AAF73728.1; -  
 DR TIGR; CP1064; -  
 DR InterPro; IPR000620; DUF6.  
 DR Pfam; PF00892; DUF6; 2.  
 SO SEQUENCE 305 AA; 33552 MW; 35820F888B0B97D1 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 942 SKLVNNA 948  
 DB 22 SKLVNNA 28

## RESULT 75

ID O56567 PRELIMINARY; PRT; 306 AA.

AC O56567;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Vibrio anguillarum virC.  
 OS Vibrio anguillarum (Listonella anguillarum).  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Listonella.  
 OX NCBI\_TaxID=55601;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RC STRAIN=NB10;  
 RA MEDLINE=96060845; PubMed=7590330;  
 RA Milton D.L., Norqvist A., Wolf-Watz H.;  
 RT "Sequence of a novel virulence-mediating gene, virC, from Vibrio  
 anguillarum.";

RL Gene 164:95-100(1995).  
 DR EMBL; U17054; AAA86984.1; -  
 SO SEQUENCE 306 AA; 34888 MW; FBC15A2A76184429 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 306;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 PVLVNNI 264  
 DB 134 PVLVNNI 140

## RESULT 76

ID O96SE2 PRELIMINARY; PRT; 309 AA.

AC O96SE2;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CGI-55 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Spiridonov N.A., Wong L., Johnson G.R.;  
 RT "Homo sapiens CGI-55 protein mRNA."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY032853; AAK51130.1; -  
 FT NON TER 1 1  
 SO SEQUENCE 309 AA; 34358 MW; C7A4CA08CE9A9E48 CRC64;

Query Match 0.7%; Score 7; DB 4; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 ADTENKE 493  
 DB 169 ADTENKE 175

## RESULT 77

ID O8Z180 PRELIMINARY; PRT; 310 AA.

AC O8Z180;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE LysR-family regulatory protein.  
 GN YPO0631.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Yersinia.  
 OX NCBI\_TaxID=632;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RC STRAIN=CO-92 / BIOVAR ORIENTALIS;  
 RA MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thompson N.R., Tittball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Baeham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis F., Dougan G.,  
 RA Feltingworth T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";

RL Nature 413:523-527(2001).  
 DR EMBL; AJ414143; CAC89485.1; -  
 DR InterPro; IPR000847; HTH\_LysR.  
 DR InterPro; IPR005119; LysR\_subst.  
 DR InterPro; IPR003662; sub\_transporter.



DR Pfam; PF00126; HTH\_1; 1.  
 DR Pfam; PF03466; LysR substrate; 1.  
 DR PRINTS; PR00039; HTHLYSR.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 310 AA; 34318 MW; EF8FB303FAC891E8 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 HUSARV 15  
 DB 277 HUSARV 283

RESULT 78  
 Q9EWZ9 PRELIMINARY; PRT; 312 AA.  
 AC Q9EWZ9; 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Purative oxidoreductase.  
 GN SC0127 OR 2SCG38.20.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyicinae; Streptomycetaceae; Streptomyces.  
 NCBI\_Taxid=1902;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Saunders D.C., Harris D.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA MEDLINE=97000351; PubMed=8843436;  
 RX Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Mulry J., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2)";  
 RL Nature 417:141-147(2002).  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 (SDR) FAMILY.  
 CC EMBL; AL445503; CAC13079.1; -  
 DR HSSP; P14061; IFDU.  
 DR InterPro; IP0002198; ADH\_short.  
 DR Pfam; PF00106; adh\_short.1.  
 DR PRINTS; PR00080; SDRFAMILY.  
 DR PROSITE; PS00061; ADH\_SHORT; UNKNOWN\_1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 312 AA; 33191 MW; C8AC48B96C0E5B73 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 312;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AGEFLDM 103  
 DB 152 AGEFLDM 158

RESULT 79  
 Q97FU1 PRELIMINARY; PRT; 315 AA.  
 AC Q97FU1; 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)  
 DE Hypothetical protein CAC2635.  
 GN CAC2635.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 NCBI\_Taxid=1489;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=21359325; PubMed=11466286;  
 RA Noelling U., Breton G., Ometchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Oiu D., Hitti J., Wolf Y.I.,  
 RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum.";  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AE007761; AAK80582.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 315 AA; 36582 MW; 1188CD3EE3A2124 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 315;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 571 KIYSINV 577  
 DB 306 KIYSINV 312

RESULT 80  
 Q98IS2 PRELIMINARY; PRT; 317 AA.  
 AC Q98IS2; 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Permease protein of oligopeptide ABC transporter.  
 GN M122275.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 NCBI\_Taxid=381;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AP002999; BAB49444.1; -

DR InterPro: IPR000515; BPD\_transp.  
 DR Pfam: PF00528; BPD\_transp.1.  
 DR PROSITE: PS00402; BPD\_TRANS\_INN\_MEMBER, UNKNOWN\_1.  
 DR Complete proteome.  
 SQ SEQUENCE 317 AA; 33247 MW; 747A2DDB9C61AEC2 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 761 SLADRLLI 767  
 |||||  
 Db 126 SLADRLLI 132

## RESULT 81

Q43055 PRELIMINARY; PRT; 318 AA.  
 AC Q43055;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Peroxidase (EC 1.11.1.7).  
 GN PRXA1.  
 OS Populus kitakamiensis (Aspen).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids I; Malpighiales; Salicaceae; Populus.  
 OC NCBI\_TaxID=3698;  
 RX MEDLINE=93169019; PubMed=7764045;  
 RA Kawai S., Matsumoto Y., Kajita K., Katayama Y.,  
 RA Morishoshi N.;  
 RT "nucleotide sequence for the genomic DNA encoding an anionic  
 RT peroxidase gene from a hybrid poplar, Populus kitakamiensis.";  
 RL Biosci. Biotechnol. Biochem. 57:131-133(1993).  
 DR EMBL: D11102; BAA01877.1; -.  
 DR HSSP: P00433; 2ATU.  
 DR InterPro: IPR002016; Peroxidase.  
 DR Pfam: PF00141; peroxidase.1.  
 DR PRINTS: PR00458; PEROXIDASE.  
 DR PROSITE: PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
 DR Oxidoreductase; Peroxidase.  
 SQ SEQUENCE 318 AA; 34172 MW; 14838148CD6076E3 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 318;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 680 VTLAGGP 686  
 |||||  
 Db 129 VTLAGGP 135

## RESULT 82

Q91E43 PRELIMINARY; PRT; 325 AA.  
 AC Q91E43;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE U2a homologue (Fragment).  
 GN U2.  
 OS Phocid herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.  
 OC NCBI\_TaxID=47418;  
 RX STRAIN=PB84;  
 RA Martina B., Osterhaus A.D.M.E., Harder T.C.;  
 SQ SEQUENCE FROM N.A.

RT "Identification and analysis of immunogenicity of the glycoprotein D  
 RT equivalent within the unique short segment of phocid herpesvirus-1.";  
 RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AJ290955; CAC51462.1; -.  
 DR InterPro: IPR003485; U2a\_Unk.  
 DR Pfam: PF02476; U2a; 1.  
 DR NCBI\_TaxID=47418;  
 SQ SEQUENCE 325 AA; 35995 MW; 2EF22B8FB483D1B9 CRC64;

Query Match 0.7%; Score 7; DB 12; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 887 LCSSGIS 893  
 |||||  
 Db 287 LCSSGIS 293

## RESULT 83

Q9CAX9 PRELIMINARY; PRT; 327 AA.  
 AC Q9CAX9;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 37.6 kDa protein.  
 GN F24K9.15.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OC NCBI\_TaxID=3702;  
 RX MEDLINE=21016720; PubMed=1130713;  
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,  
 RA Fattmann B., Valle G., Bioeche H., Perez-Alonso M., Obermaier B.,  
 RA Deisenhoef V., Cholewa-Waclaw J., Mach R., Puigdomenech P.,  
 RA Wincker P., Catolico L., Weissenbach J., Saurin W., Queller F.,  
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Wurmbach E., Drzonek H., Erle H., Jordan N., Bangert S.,  
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
 RA Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T., Nordstiek B.,  
 RA Reichelt J., Scharte M., Schoen O., Bargues M., Terol J., Clement J.,  
 RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,  
 RA Cooke R., Laudie M., Berger-Hauro C., Puntel A., Casasuberta E.,  
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casasuberta E.,  
 RA Monfort A., Argitrou A., Flores M., Liguori R., Vitale D.,  
 RA Mannhaupt G., Haase D., Schoof H., Ruid S., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Walts A., Ulteback T., Fujii C.Y., Shea T.P.,  
 RA Creasy T.H., Haas B., Maltz R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Miltscher J., Sellers P., Gill J.E., Feldblum T.V.,  
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Frazer C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 RA Saenmoko S., Kimura T., Igesawa K., Kawashima K., Kishida A.,  
 RA Nakayama S., Kohara M., Matsumoto M., Matsumoto K., Muraki A.,  
 RA Kikuyama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.,  
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:820-822(2000).  
 DR EMBL: AC008153; AAG51446.1; -.  
 DR InterPro: IPR005299; Methyltransf\_6.  
 DR Pfam: PF03492; Methyltransf\_6; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 327 AA; 37575 MW; 8B3FD25A11B8D700 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 257 KEVLVYN 263

Db 3 KEVLVYN 9

## RESULT 84

09POG1 PRELIMINARY; PRT; 327 AA.

AC 09POG1; DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

GN Hypothetical protein U0330.

OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmataceae; Ureaplasma.

OX NCBI\_TaxID=134821;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SERVAR 3; PubMed=11048724;

RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,

RA Cassell G.H.; "The complete sequence of the mucosal pathogen Ureaplasma

urealyticum.";

RL Nature 407:757-762(2000).

DR EMBL; AB002130; AAF30739.1; "

KM Hypothetical protein; Complete proteome.

SO SEQUENCE 327 AA; 38114 MW; 321275386DAF61B7 CRC64;

Query Match Best Local Similarity 0.7%; Score 7; DB 16; Length 327;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 469 FMILTV 475

Db 170 FMILTV 176

## RESULT 85

097KX8 PRELIMINARY; PRT; 336 AA.

AC 097KX8; DT 01-OCT-2001 (TREMBLrel. 18, Created)

DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

GN Perichrome transport permease.

OS CAC0788.

OC Clostridium acetobutylicum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Clostridiaceae; Clostridium.

OX NCBI\_TaxID=1488;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

RX MEDLINE=21359325; PubMed=11466286;

RA Neilling J., Berton G., Omeichenko M.V., Makarova K.S., Zeng Q.,

RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,

RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,

RT "Genome sequence and comparative analysis of the solvent-producing

bacterium Clostridium acetobutylicum.";

RL J. Bacteriol. 183:4823-4838(2001).

DR EMBL; AB007594; AAK78764.1; "

DR InterPro; IPR000522; FecCD.

DR Pfam; PF01032; FecCD; 1.

DR ProDom; PD001557; FecCD; 1.

KM Complete proteome.

SO SEQUENCE 336 AA; 36552 MW; 1BC0BC10C47B0F34 CRC64;

Query Match Best Local Similarity 0.7%; Score 7; DB 16; Length 336;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 921 ILTVLT 927

Db 19 ILTVLT 25

## RESULT 86

08RVP3 PRELIMINARY; PRT; 347 AA.

AC 08RVP3; DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

GN Bacterial-induced peroxidase (EC 1.11.1.7).

OS Gossypium hirsutum (Upland cotton).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Malvales; Malvaceae; Gossypium.

OX NCBI\_TaxID=3635;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. REBA B50; TISSUE=2 WEEK OLD COTYLEDONS;

RA Delannoy E., Patil M., Essenberg M., Jalloul A., Marney P.,

RA Daniel J.-F., Geiger J.-P., Nicole M.;

RT "Gossypium hirsutum bacterial-induced peroxidase (podSSH1).";

RL Submitted (FBI-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF48305; AAL92037.1; "

KM Peroxidase; Oxidoreductase.

SO SEQUENCE 347 AA; 37407 MW; 51B68CA3CCF7F0BB CRC64;

Query Match Best Local Similarity 0.7%; Score 7; DB 10; Length 347;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 681 TLGGPS 687

Db 137 TLGGPS 143

## RESULT 87

08ZOT2 PRELIMINARY; PRT; 350 AA.

AC 08ZOT2; DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

GN 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase (DAHPSynthetase,

DAHPSynthetase).

OS Salmomella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmomella.

OX NCBI\_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=L75 / SCS1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Potwolk S., Ali J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.; "Complete genome sequence of Salmomella enterica serovar Typhimurium

LT2.";

RL Nature 413:852-856(2001).

DR EMBL; AB008731; AAL19699.1; "

DR InterPro; IPR001785; DAHP synth.

DR Pfam; PF00793; DAHP synth.1; 1.

DR ProDom; PD005060; DAHP synth.1.

DR TIGR0034; atofGH; 1.  
 KW Lyase; Complete proteome.  
 SQ SEQUENCE 350 AA; 37856 MW; D79AC095B456A147 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 AGEFLDM 103  
 |||||  
 Db 141 AGEFLDM 147

RESULT 88  
 Q828B5 PRELIMINARY; PRT; 350 AA.

AC Q828B5;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DE Phospho-2-dehydro-3-deoxyheptanate aldolase (DAH synthetase)  
 DE phenylalanine repressible.  
 GN STY0801.

OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gara P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrett B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18";  
 RL Nature 413:848-852(2001).  
 DR EMBL: AL627268; CAD05217.1; -;  
 DR InterPro: IPR001785; DAHP\_synth1.  
 DR Pfam: PF00793; DAHP\_synth\_1; 1.  
 DR ProDom: PD005060; DAHP\_synth1; 1.  
 DR TIGR0034; atofGH; 1.  
 KW Complete proteome.

SQ SEQUENCE 350 AA; 37868 MW; C764207BBA65FB7 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 AGEFLDM 103  
 |||||  
 Db 141 AGEFLDM 147

RESULT 89  
 Q85Y1 PRELIMINARY; PRT; 357 AA.

AC Q85Y1;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical protein ECU07\_1170.  
 GN ECU07\_1170.

OC Encephalitozoon cuniculi.  
 OC Eukaryote; Microsporidia; Unikaryonidae; Encephalitozoon.  
 OX NCBI\_TaxID=6035;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=GB-M1;  
 RA Genoscope; (APR-2001) to the EMBL/Genbank/DBJ databases.  
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=GB-M1;  
 RX MEDLINE=21576510; PubMed=11719806;  
 RA Katinka M.D., Duprat S., Cornille E., Metenier G., Thomarat F.,  
 RA Prensier G., Barbe V., Peyretallade E., Broctier P., Winkler P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissenbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT Encephalitozoon cuniculi";  
 RL Nature 414:450-453(2001).  
 DR EMBL: AL590447; CAD25650.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 357 AA; 40257 MW; 715D55ED5BC2414 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 663 NTPRTF 669  
 |||||  
 Db 166 NTPRTF 172

RESULT 90  
 Q92HR6 PRELIMINARY; PRT; 360 AA.

AC Q92HR6;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Similar to microcin C7 self-immunity protein.  
 GN MCCP2 OR RC0705.

OS Rickettsia conorii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=781;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MALISH 7;  
 RX MEDLINE=21442074; PubMed=11557893;

RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";  
 RL Science 293:2093-2098(2001).  
 DR EMBL: AE008629; AAL03243.1; -;  
 DR InterPro: IPR003507; UPF0094.  
 DR Pfam: PF02016; UPF0094; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 360 AA; 40394 MW; B94412DFC9831C95 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 874 G10KTY 880  
 |||||  
 Db 5 G10KTY 11

RESULT 91  
 Q9CN76 PRELIMINARY; PRT; 361 AA.

AC Q9CN76;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE AroG.  
 GN AROG OR PM0563.

OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 OX NCBI\_TaxID=147;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.,  
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).  
 DR EMBL; AB06092; AAK02647.1; -.  
 DR HSSP; P00886; 1QR7.  
 DR InterPro; IPR001785; DAHP\_synth1.  
 DR Pfam; PF00783; DAHP\_synth\_1; 1.  
 DR Prodom; PD005060; DAHP\_synth; 1.  
 DR TIGRFAMs; TIGR00034; aroFGH; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 361 AA; 40132 MW; 4B78F085186C1B2E CRC64;

Query Match 0.7%; Score 7; DB 16; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AGEFLDM 103  
 Db 148 AGEFLDM 154

RESULT 92  
 ID 095KAO PRELIMINARY; PRT; 364 AA.  
 AC 095KAO;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical 39.3 kDa protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MEDULLA OBLONGATA;  
 RA Osada N., Hida M., Kuwada J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RA Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 RT libraries."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB063042; BAB60776.1; -.  
 DR InterPro; IPR000873; AMP-bind.  
 DR Pfam; PF00501; AMP-binding; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 364 AA; 39316 MW; 46F7998483BFE746 CRC64;

Query Match 0.7%; Score 7; DB 6; Length 364;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 PAGTEPA 418  
 Db 350 PAGTEPA 356

RESULT 93  
 ID 09FC16 PRELIMINARY; PRT; 364 AA.  
 AC 09FC16;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative secreted protein.

GN SC05202 OR 28C3B6.26.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutter S., Sanger K., Saunders D.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL390668; CAC01332.1; -.  
 DR InterPro; IPR001478; PDZ.  
 DR SMART; SM00228; PDZ; 1.  
 SQ SEQUENCE 364 AA; 38181 MW; 8900C095D03C17B1 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 364;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 911 VGISAGT 917  
 Db 232 VGISAGT 238

RESULT 94  
 ID P91984 PRELIMINARY; PRT; 371 AA.  
 AC P91984;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE C50B6.10 protein.  
 GN C50B6.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Percy C.M.;  
 RT Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;

RA none;  
 RT "genome sequence of the nematode C.elegans: A platform for  
 RL investigating biology."; 1998.  
 DR Science 282:2012-2018(1998).  
 DR EMBL; Z81050; CAB02859.1; -.  
 DR InterPro; IPR003002; 7TM\_chemol.  
 DR InterPro; IPR000168; 7TM\_nematode.  
 DR Pfam; PF01461; 7tm\_4; 1.  
 SQ SEQUENCE 371 AA; 42447 MW; 872EEC33780B436D CRC64;

Query Match 0.7%; Score 7; DB 5; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 VGVNSRT 521  
 DB 360 VGVNSRT 366

RESULT 95  
 Q9CBV6 PRELIMINARY; PRT; 371 AA.  
 AC Q9CBV6;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE l-alanine dehydrogenase.  
 GN ALD OR ML1532.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterinae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_Taxid=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Sigmeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,  
 RA Holtz S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrett B.G.;  
 RT "Massive gene decay in the leprosy bacillus";  
 RL Nature 409:1007-1011(2001).  
 CC -i- COFACTOR: FAD (BY SIMILARITY).  
 DR EMBL; AL583922; CAC30483.1; -.  
 DR HSSP; Q60164; 1H2Z.  
 DR Leproma; ML1532; -.  
 DR InterPro; IPR004002; ALADH\_PNT.  
 DR InterPro; IPR001327; FAD\_PNT\_redox.  
 DR InterPro; IPR000205; NAD\_binding.  
 DR Pfam; PF01262; ALADH\_PNT; 1.  
 DR PRINTS; PRO0368; FADPNT.  
 DR TIGRFAMs; TIGR00518; aladh; 1.  
 DR PROSITE; PS00836; ALADH\_PNT\_1; 1.  
 DR PROSITE; PS00837; ALADH\_PNT\_2; 1.  
 DR FAD; Flavoprotein; Oxidoreductase; Complete proteome.  
 SQ SEQUENCE 371 AA; 38845 MW; 35425DE604EF8199 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 DLGAVK 369  
 DB 224 DLGAVK 230

RESULT 96  
 Q9CRR7

ID Q9CRR7 PRELIMINARY; PRT; 377 AA.  
 AC Q9CRR7;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE MAD homolog 5 (Drosophila) (Fragment).  
 GN MADH5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=EMERSONIC HEAD;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukushima Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Holtman M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK018077; BAB31063.1; -.  
 DR HSSP; Q92940; 1MDH.  
 DR MGD; MGI:1328787; Madh5.  
 DR InterPro; IPR001132; Dwarf1n.  
 DR InterPro; IPR003619; Dwarf1n\_A.  
 DR Pfam; PF03165; MH1.  
 DR Pfam; PF03165; MH1; 1.  
 DR Pfam; PF03166; MH2; 1.  
 DR SMART; SM00523; DWA; 1.  
 DR SMART; SM00524; DWA; 1.  
 FT NON\_TER  
 SQ SEQUENCE 377 AA; 42204 MW; 830E5E1452D025DE CRC64;

Query Match 0.7%; Score 7; DB 11; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 FHSVELN 236  
 DB 339 FHSVELN 345

RESULT 97  
 Q91WP3 PRELIMINARY; PRT; 386 AA.  
 AC Q91WP3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE Similar to RIKEN CDNA 1200009K13 gene.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SALIVARY GLAND;  
 RA Strausberg R.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC013665; AAH13665.1; -  
 SQ SEQUENCE 386 AA; 42235 MW; A22PB50BE68F72CB CRC64;  
 Query Match 0.7%; Score 7; DB 11; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 487 ADTENKE 493  
 |||||  
 Db 246 ADTENKE 252

RESULT 98  
 ID 069648 PRELIMINARY; PRT; 386 AA.  
 AC 069648;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DE Putative anion transporting ATPase (Hypothetical protein MT3782).  
 GN RV3680 OR MT025.028 OR MT3782.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NC NCB1\_TaxID=1773;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RC MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,  
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Stulton J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / OSHKOSH;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M., Hatt D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains." to the EMBL/GenBank/DDBJ databases.  
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL022121; CAI18002.1; -  
 DR EMBL; AE007175; AAK48149.1; -  
 DR TIGR; MT3782; -  
 DR TubercuList; RV3680; -  
 DR InterPro; IPR003348; Arsa\_ATPase.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 386 AA; 41404 MW; ABED2031A44620B5 CRC64;

Query Match 0.7%; Score 7; DB 16; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 24 RLMRLIL 30  
 |||||  
 Db 178 RLMRLIL 184

RESULT 99  
 QY483 PRELIMINARY; PRT; 387 AA.  
 ID QY483

AC QY483;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Hypothetical 42.4 kDa protein (PAI-1 mRNA-binding protein).  
 GN DKFZ564M2423.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRNIN;  
 RC MEDLINE=21154917; PubMed=11230166;  
 RA Wieman S., Weil B., Wellenreuther R., Gassenhuber J., Glaser S.,  
 RA Ansoerge W., Boecker M., Bloeker H., Baerach S., Blum H.,  
 RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
 RA Mewes H.W., Oettermeyer B., Obermayer B., Tampe J., Heubner D.,  
 RA Wambutt R., Korn B., Klein M., Poustka A.;  
 RT "Towards a Catalog of Human Genes and Proteins: Sequencing and  
 RT Analysis of 500 Novel Complete Protein Coding Human cDNAs."  
 RL Genome Res. 11:422-435(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RA Strausberg R.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL080119; CAB45718.1; -  
 DR EMBL; BC017449; AAH17449.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 387 AA; 42426 MW; 8C47134D22C1CCFA CRC64;

Query Match 0.7%; Score 7; DB 4; Length 387;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 487 ADTENKE 493  
 |||||  
 Db 247 ADTENKE 253

RESULT 100  
 QY9PC12  
 ID 09PC12 PRELIMINARY; PRT; 389 AA.  
 AC 09PC12;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Porphyryrin biosynthesis protein.  
 GN XP1797.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 NC NCB1\_TaxID=2371;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RC MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Keinich F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Britton M.R.S.,  
 RA Bueno M.R.P., Colombo A.A., Camargo L.E.A., Carraro D.M., Carter H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Pachican A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 RA Gartner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitejima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miranda E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Mnaní A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pequeiro J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldanis J., Zubal J.C.,  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE004001; AAF84605.1; -.  
 DR InterPro; IPR001440; TPR.  
 KM Complete proteome.  
 SQ SEQUENCE 389 AA; 44436 MW; 2459773D3632925 CRC64;

## Query Match

0.7%; Score 7; DB 16; Length 389;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 988 LPPRGLL 994  
 |||||  
 DB 181 LPPRGLL 187

Search completed: April 22, 2003, 15:35:32  
 Job time : 72 secs